INDEX

Preface

As you will have noticed, NAKE is using the electronic highway more and more. For example, last year, we started to do the evaluation of the Utrecht courses via email and recently we introduced an internet questionnaire to investigate the students' preferences for courses next year. Although these innovations were succesful (response was much higher than it used to be), we will not completely abolish communication via snail mail. In particular, we will stick to the tradition of sending you three issues of *NAKE Nieuws* per year.

In this issue, you will find the last report on the workshop that was held in Groningen, June 11-15, 2000. **Jeroen Kerkhof** (KUB) writes on "The Bootstrap" by Joel Horowitz. Our most recent workshop was held December 11-15, 2000 in Maastricht. Although the number of participants was a bit low, it was a very succesful week. Flying Dutchman **Guido Imbens** gave a series of lucid lectures on "Causal Inference and Program Evaluation". **Kaushik Basu** demonstrated in his series of lectures on "Rationality and Social Norms" that he is not only a great scientist but also a very good storyteller. **Hyun Song Shin** presented a number of very well structured lectures providing a better insight in the theory and models of financial crises. And last but not least, the series of smooth Power-Point presentations by **Gerard Roland** corroborated his reputation as one of the foremost specialists in the field of "Transition and Economics". Only the reports on the latter two series of lectures are included in this *NAKE Nieuws*. **Mindel van de Laar** (UM) wrote a very concise report on the extensive material covered by Roland. **Remco van der Molen** (RUG) lucidly surveyed Shin's lectures. For the reports on Basu and Imbens you will have to wait for the next issue.

Life does go on of course. The next workshop will be at the Vrije Universiteit Amsterdam. **Peter Howitt** (Brown University) will lecture on "Innovation-based Growth Theory". **Preston McAfee** (University of Texas at Austin) will give the microeconomics lectures on the topic of "Pricing". The econometrics lectures will be given by **Ken Judd** (Hoover Institution, Stanford University). He will talk about "Numerical Methods for Dynamic Economic Analysis". Finally Kai Konrad, (Free University of Berlin) will lecture on "Contests: Theory and Applications". Elsewhere in this *NAKE Nieuws* you find details on the programme. You can still register for the workshop via the NAKE homepage.

Currently, NAKE is in the middle of the process of planning the courses for next year. The courses for the first block (September 7 - October 19) are already known (see Upcoming... on page 58). At the moment, we are working on the second block. The number of courses in this block will be a bit higher than usual. The reason for this is that we received signals that there is a considerable interest in PhD courses on (empirical) finance. As NAKE has fellows with great expertise in this field, we decided to offer a number of courses on this topic in addition to the 'standard' courses. In order to make it worthwhile for students especially interested in finance to travel to Utrecht, we concentrated these courses in one block instead of spreading them over four blocks. We plan to organise more of these sub-blocks concentrated on a specific subject on the borderline between economics and business. For example, we are currently investigating the possibilities for courses on the economic theory of organisation. The complete course programme for next year will soon be on our homepage. There you can also register for the courses. I would like to urge you to timely enrol for courses you want to follow as we will cancel courses for which a week before the start of the block less than 5 participants have registered. Also, please inform the NAKE secretariat if you decide not (or no longer) to follow a course for which you have enrolled.

Finally, I would like to already mention that the next NAKE Research Day is planned on Friday October 12, 2001. Please reserve this important date in your diary now.

Best regards,

Lex Meijdam

NAKE WORKSHOP

11 - 15 June 2001

Vrije Universiteit Amsterdam (VU)

During the week from Monday 11 June to Friday 15 June 2001 the Netherlands Network of Economics (NAKE) will organize a PhD. workshop. Four distinguised economists will teach intensive courses on micreconomics, macroeconomics, econometrics and public economics. Each course consists of five lectures spread out over five days.

Courses

Peter Howitt (Brown University)
"Innovation-based Growth Theory"

Ken Judd (Hoover Institution, Stanford University)
"Numerical Methods for Dynamic Economic Analysis"

Kai Konrad (Free University of Berlin)
"Contests: Theory and Applications"

Preston McAfee (University of Texas at Austin)
"Pricing"

For more information see: http://few.kub.nl/nake/forthcomingws.htm

Teaching programme NAKE workshop - VU - 11-15 June 2001

Monday June 11

10.00 -	11.00	registration/coffee
11.00 -	12.15	Konrad
12.15 -	**13.30**	**Lunch**
13.30 -	14.45	McAfee
15.00 -	16.15	Judd
16.30 -	17.45	Howitt
17.45 -	**19.15**	**Welcome reception**

Tuesday June 12

09.00 -	10.45	Konrad
11 .00 -	12.45	Judd
12.45 -	**14.00**	**Lunch**
14.00 -	15.45	Howitt
16.00 -	17.30	Konrad

Wednesday June 13

09.00 -	10.30	McAfee
10.45 -	12.15	Howitt
12.15 -	**13.30**	**Lunch**
13.30 -	15.00	Judd
15.15 -	17.00	McAfee
17.00 -	**18.30**	**Individual consultations**

Thursday June 14

09.00 -	10.45	Judd
11.00 -	12.45	Konrad
12.45 -	**14.00**	**Lunch**
14.00 -	15.45	McAfee
16.00 -	17.45	Howitt
20.00		**Workshop dinner**

Friday June 15

09.00 -	10.30	McAfee
10.45 -	12.15	Judd
12.15 -	**13.30**	**Lunch**
13.30 -	15.00	Howitt
15.15 -	16.45	Konrad
16.45 -	**...**	**Closing drinks**

Note: All lectures except those on Monday include a short (5-10 minutes) break.

The bootstrap
Joel Horowitz

Report by Jeroen Kerkhof, Tilburg University[*]

Abstract

In this paper, we investigate the bootstrap. This is a statistical technique to calculate the distribution of statistics (e.g. estimators or test statistics). We derive and exemplify the possibilities and virtues of the bootstrap.

1 Introduction

A well-known problem in econometrics is finding the distribution of an estimator or test statistic in a finite sample. Usually, applied econometricians rely on first-order asymptotic theory to solve this problem. This can be quite inaccurate, especially in small samples. In this paper we therefore investigate the bootstrap, which is an alternative method for estimating the distribution of an estimator or test statistic in a finite sample by resampling the data. The main idea of the bootstrap is treating the data as if they were the whole population. As we show in the remainder, the bootstrap has two important advantages compared to first-order asymptotic theory. First, it substitutes the often difficult mathematical analysis of first-order asymptotic theory by computer power. Second, under some mild regularity conditions the approximation obtained by the bootstrap is at least as good as that obtained by first-order asymptotic theory. The set-up of this paper is as follows. We start with the most well-known model in econometrics, the classical regression model to get a flavour for the possibilities of the bootstrap. In Section 3; we discuss the general problem. Subsequently, the theory behind the bootstrap is explained. We discuss some selected topics. Finally, we illustrate the presented theory with simulation studies and conclude.

---

[*]Correspondence to: Jeroen Kerkhof, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. I would like to thank Laurens Swinkels for the various discussions to enlarge our understanding of the bootstrap.

2 Motivating example: Classical linear regression

The most often used model in econometrics is without any doubt the classical regression model. It therefore seems a good starting point to illustrate the possible virtues of the bootstrap within this model. Recollecting the set-up of the classical regression model. We have available a data set containing $\{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\}$. We assume that the data are a realization of a random sample $\{(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$ from an unknown distribution F_0. Our model is the classical regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{1}$$

with an i.i.d. sequence $\{\varepsilon_i\}$ such that

$$\begin{aligned} \mathbb{E}(\varepsilon_i) &= 0 \\ \mathbb{E}(X_i \varepsilon_i) &= 0 \\ \mathbb{E}\left(\varepsilon_i^2\right) &= \sigma^2. \end{aligned} \tag{2}$$

We estimate (β_0, β_1) by the OLS estimator b_{OLS}[1]

$$b_{OLS} = (X'X)^{-1} X'Y \tag{3}$$

where

$$X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \text{ and } Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

and σ^2 by s^2 [2]

$$s^2 = \frac{1}{n-2}(Y - Xb)'(Y - Xb) \tag{4}$$

We would like to know whether X_i helps in explaining Y_i. Stated more formally, we would like to test

$$H_0 : \beta_1 = 0 \text{ vs. } H_0 : \beta_1 \neq 0 \tag{5}$$

To perform this test, we use a test statistic $T_n = T_n(X_1, \ldots, X_n)$:

$$T_n(X_1, \ldots, X_n) = \sqrt{n}\frac{b_1}{s\,(X'X)^{-1}_{22}} \tag{6}$$

---

[1] We also denote the OLS estimate $b_{OLS} = (x'x)^{-1} x'y$ by b_{OLS}. where x, y denote realizations of X, Y, respectively. It will be clear from the context which is meant.

[2] Here also s^2 denotes both the estimator and the estimate for σ^2.

If we assume that $\mathcal{L}(\varepsilon_i) = N(0; \sigma^2)$[3], we have that under some regularity conditions

$$\mathcal{L}(b_{OLS}) = N\left(\beta; \operatorname{plim} \sigma^2 (X'X)^{-1}\right)$$

Standard statistical books give that

$$\mathcal{L}(T_n) = t_{n-2}.$$

We have found the distribution for the test statistic T_n in a finite sample. To achieve this result, we had to make distributional assumptions on the errors, ε_i. Often, we are not too sure about the distribution of the errors and would like to relax this assumption to the assumptions in (2). The traditional approach is to use asymptotic theory. It is well known that if n is large,

$$\mathcal{L}(b_{OLS} | X = x) \approx N\left(\beta; s^2 (x'x)^{-1}\right).$$

If we are willing to make the assumption that our sample is large enough, we find

$$\mathcal{L}(T_n) \approx N(0; 1)$$

and can base inference on this distribution. But how do we know when n is large enough? There is no clear-cut answer to this question. Usually, we design our test statistic such that it rejects a correct null hypothesis with probability α[4]. Since, the distribution of the test statistic is only approximate, it could be the case that a true null hypothesis is rejected far more often or less than wanted. Therefore, we would like to get a better approximation of $\mathcal{L}(T_n)$. Suppose, that instead of just one realization of $\{(Y_1; X_1); (Y_2; X_2); \ldots; (Y_n; X_n)\}$ we have N. This gives us N realizations of the test statistic as well. Now, if N is large, we get a good approximation of the distribution of T_n. It would be therefore be very useful to have more realizations of $\{(Y_1; X_1); (Y_2; X_2); \ldots; (Y_n; X_n)\}$. Since F_0 is unknown, we cannot draw from it. What we can do is draw n times from our initial data with replacement and create artificial realizations of size n. This is basically the bootstrap. In Section 6.2, we see that in this example the bootstrap provides us with a better approximation of the finite sample distribution of the test statistic than asymptotic theory. Before we get there, we first look at the general problem trying to understand why the bootstrap provides us with such a better approximation. In the next section, we take a look at the more general problem.

---

[3] $\mathcal{L}(X)$ denotes the law (or probability distribution) of X.

[4] $\alpha = 0.05$ is most often used.

3 General problem

Assume that the data set $\{x_1, x_2, \ldots, x_n\}$ that we have available is a realization of a random sample $\{X_1, X_2, \ldots, X_n\}$ from a probability distribution with cumulative distribution function (CDF) F_0. F_0 may belong to a finite- or infinite-dimensional family $\mathcal{F}$. If F_0 belongs to a finite-dimensional family $\mathcal{F}(\theta)$ indexed by a parameter θ whose population value is θ_0, we write $F_0(x; \theta_0) = \mathbb{P}_{\theta_0}(X_i \leq x)$[5] and $F(x; \theta) = \mathbb{P}_\theta(X_i \leq x)$ for a general member of $\mathcal{F}(\theta)$. Let $T_n = T_n(X_1, X_2, \ldots, X_n)$ denote the test statistic of interest for testing a hypothesis about a population parameter θ or a function $r(\theta)$ of θ. Let $G_n(t; F_0) \equiv \mathbb{P}_{\theta_0}(T_n \leq t)$ denote the exact finite-sample CDF of T_n under the null hypothesis H_0.

The α-level critical $z_{n\alpha}$ value solves

$$G_n(z_{n\alpha}; F) = 1 - \alpha \tag{7}$$

in the case of a one-sided test and

$$G_n(z_{n\alpha}; F) - G_n(-z_{n\alpha}; F) = 1 - \alpha \tag{8}$$

in the case of a two-sided test. In the remainder of the paper, we concentrate on the one-sided test, since this illustrates the virtues of the bootstrap without loss of generality. Furthermore, the theory of the two-sided test is easily recovered from the theory of the one-sided test.

In most applications, $G_n(\cdot; F)$ depends on F, which is unknown, and therefore (7) cannot be solved. In some special cases, $G_n(\cdot; F)$ does not depend on F, in which case we call T_n a pivotal statistic. A well-known example is the usual t-statistic corresponding to the estimate of a mean from a standard normal distribution. F is a normal distribution and from this is it follows that G_n has a t-distribution with $n - 1$ degrees of freedom. Pivots typically exist under strong distributional assumptions (to be more precise, you have to make distributional assumptions in order to obtain a pivot). If T_n is not pivotal (the usual case), we cannot solve (7). Therefore, we have to find an approximation for $z_{n\alpha}$. One way to handle the problem of approximating the critical values of the test-statistic T_n is resorting to asymptotic distribution theory. An alternative approach, the bootstrap, has been introduced by Efron (1979). First, we discuss the traditional approach using asymptotic theory. Second, we investigate the bootstrap. Finally, we compare the two methods.

---

[5] When it is clear from the context which measure is used, we will supress the subscripts.

3.1 Asymptotic distribution theory

Though few statistics are pivotal, most statistics used in econometrics are asymptotically pivotal. A statistic is asymptotically pivotal if its asymptotic distribution does not depend on unknown population parameters. This distribution is often the standard normal distribution or χ^2_2 distribution. More formally, let

$$G_\infty(t;F) \equiv \lim_{n\to\infty} G_n(t;F). \tag{9}$$

If T_n is asymptotically pivotal,we have

$$G_\infty(t;F) = G_\infty(t). \tag{10}$$

If n is sufficiently large, $G_\infty(t;F)$ is arbitrarily close to $G_n(t;F)$ as can be seen from (9). Furthermore, (10) states that G_∞ does not depend on F if T_n is asymptotically pivotal, so the unknown $G_n(t;F)$ can be approximated by $G_\infty(t)$. We can now get approximate solutions for (7) by replacing it with

$$G_\infty(z_{\infty\alpha}) = 1 - \alpha. \tag{11}$$

For a given level α, $G_\infty(\cdot)$ can be inverted and thereby the asymptotic critical value $z_{\infty\alpha}$ can be recovered to serve as an approximation of $z_{n\alpha}$.

3.2 The bootstrap

The bootstrap provides an alternative approximation to the finite-sample distribution of a test statistic T_n. Where first-order asymptotic approximations replaces the unknown distribution function G_n by G_∞ the bootstrap replaces the unknown CDF F_0 with a known estimator which we denote F_n. Two possible estimators F_n are:

- The empirical distribution function (EDF) of the data:

$$F_n(x) = \frac{1}{n}\sum_{i=1} \mathbf{I}_{(-\infty;x)}(X_i) \tag{12}$$

where $\mathbf{I}_{(A)}(x)$ denotes an indicator function[6].

[6] $\mathbf{I}_{(A)}(x) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \end{cases}$

- A parametric estimator of F_0: If $F_0(\cdot) = F(\cdot;\theta_0)$ for some finite dimensional θ_0, consistently estimated by θ_n; and $F(\cdot;\theta)$ is a continuous function of θ in the neighborhood of θ_0; then $F(x;\theta_n) \to F(x;\theta_0)$ for all x as $n \to \infty$.

Regardless of the choice of F_n the bootstrap estimator of $G_n(\cdot;F_0)$ is $G_n(\cdot;F_n)$. Usually, $G_n(\cdot;F_n)$ also cannot be evaluated analytically. Using a Monte-Carlo simulation drawing samples from F_n we can, however, estimate it with arbitrary accuracy. Therefore, the bootstrap is usually implemented using a Monte-Carlo simulation. The essential characteristic of the bootstrap is, however, the use of F_n to approximate F_0 in $G_n(\cdot;F_0)$, not the Monte-Carlo method used to evaluate $G_n(\cdot;F_n)$.

The Monte-Carlo procedure for estimating $G_n(t;F_n)$ can be implemented in the following way:

1. Generate L bootstrap samples of size n by sampling the distribution corresponding to F_n randomly. Denote the realizations of the random samples $\{\{X_1^*, X_2^*, \ldots, X_n^*\}_j;\ j = 1, \ldots, L\}$ by $\{\{x_1^*, x_2^*, \ldots, x_n^*\}_j;\ j = 1, \ldots, L\}$. If F_n is the EDF of the estimation data set, then the bootstrap sample can be obtained by sampling the estimation data with replacement.
2. Compute the statistics $\left\{T_{n,j}^* \equiv T_{n,j}(x_1^*, x_2^*, \ldots, x_n^*);\ j = 1, \ldots, L\right\}$.
3. Use the result of many replications of step 1 and 2 to compute the EDF of T_n^* (e.g. the proportion of T_n^* values smaller than t).

The bootstrap version of (7) is

$$G_n(z_{n\alpha}^*;F_n) = 1 - \alpha. \tag{13}$$

Inverting the EDF, F_n, of T_n^* we get the bootstrap critical value $z_{n\alpha}^*$ which serves as the bootstrap approximation for $z_{n\alpha}$.

3.3 Asymptotic theory vs the bootstrap

Recollecting the results of the two subsections above; we found two different approximations of the unknown distribution $G_n(t;F_0)$:

1. $G_\infty(t;F_0)$ by using asymptotic theory.

2. $G_n(t; F_n)$ by using the bootstrap.

This gives rise to two natural questions:

- What is the relation between $G_\infty(t; F_0)$ and $G_n(t; F_n)$?
- What is the relation between $G_\infty(t; F_0)$, $G_n(t; F_n)$ and $G_n(t; F_0)$?

For n sufficiently large $G_\infty(t; F_0)$ and $G_n(t; F_0)$ will be close. For the bootstrap, we need in addition n to be large in order for F_n to be close to F. If $G_n(\cdot; F_n)$ and $G_\infty(\cdot; F_n)$ are "continuous" functions in their second arguments near F, we expect that $G_n(t; F_n) \to G_\infty(t; F_0)$ a.s. as $n \to \infty$. This preferable property is roughly speaking consistency of the bootstrap. In the next section, consistency is defined more precise.

4 Theory of the bootstrap

In the previous section, we discussed the main idea of the bootstrap and its implementation. In this section, we become a bit more formal and discuss the theory underlying the bootstrap. We start by defining the concept of consistency, which was already informally touched upon in the last section. If we know that the bootstrap is consistent, we can apply it and thereby avoid the often complicated first-order asymptotic mathematics. This is the first advantage of the bootstrap. Furthermore, we investigate the accuracy of the bootstrap. We show that using the bootstrap we can get results that are better than using first-order asymptotic theory. This shows the second advantage of the bootstrap.

4.1 Consistency of the bootstrap

Since $G_n(t; F_0)$ is the distribution of interest and we use $G_n(t; F_n)$ as an approximation for it, we are interested in the difference $|G_n(t; F_0) - G_n(t; F_n)|$. A natural requirement is to have that $|G_n(t; F_0) - G_n(t; F_n)|$ is small if n is large. In order to investigate this, we need a metric (measure of distance) and a limiting concept as $n \to \infty$. As metric for the "distance" between two distributions G_1 and G_2, we use

$$d(G_1; G_2) \equiv \sup_t |G_1(t) - G_2(t)|$$

So, the distance between $G_n(t; F_n)$ and $G_\infty(t; F)$ is defined as $\sup_t |G_n(t; F_n) - G_\infty(t; F)|$. We can now define consistency.

Definition 1 Let $\mathbb{P}_n$ denote the joint distribution function of a random sample $\{X_1, X_2, \ldots, X_n\}$. The bootstrap estimator $G_n(\cdot, F_n)$ is said to be consistent if for every $\varepsilon > 0$ and each $F_0 \in \mathcal{F}$

$$\lim_{n \to \infty} \mathbb{P}_n(\sup_t |G_n(t, F_n) - G_\infty(t, F)| > \varepsilon) = 0 \tag{14}$$

The metric for the space of distributions $\rho_n = d(F_n, F_0)$ is the "distance" between F_n and F_0.

It is rather difficult to check consistency straight from the definition. A useful theorem with sufficient conditions for consistency is provided by Beran and Ducharme (1991).

Theorem 2 $G_n(\cdot, F_n)$ is consistent if for any $\varepsilon > 0$ and $F_0 \in \mathcal{F}$:

1. $\mathbb{P}_n(\rho(F_n, F_0) > \varepsilon) \to 0$ as $n \to \infty$ (note: condition on F and not G)
2. $G_\infty(t, F)$ is a continuous function of t for each F.
3. For any t and any sequence of distribution functions H_n such that $\rho(H_n, F) \to 0$ and $G_n(t, H_n) \to G_\infty(t, F)$ as $n \to \infty$. This condition is called the "triangular array convergence condition".

The theorem by Beran and Ducharme (1991) gives sufficient conditions for consistency, not necessary conditions. The next theorem by Mammen (1992) gives necessary and sufficient conditions for consistency, but for a smaller class of test statistics, namely the class of linear test statistics.

Theorem 3 Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a population. For a sequence of functions g_n and sequences of numbers t_n and σ_n, define $\bar{g}_n = \frac{1}{n}\sum_{i=1} g_n(X_i)$ and $T_n = (\bar{g}_n - t_n)/\sigma_n$. For the bootstrap sample $\{X_1^*, X_2^*, \ldots, X_n^*\}$ define $\bar{g}_n^* = \frac{1}{n}\sum_{i=1} g_n(X_i^*)$ and $T_n^* = (\bar{g}_n^* - \bar{g}_n)/\sigma_n$. Let $G_n(\tau) = \mathbb{P}(T_n \leq \tau)$ and $G_n^*(\tau) = \mathbb{P}^*(T_n^* \leq \tau)$. Where $\mathbb{P}^*$ is the probability distribution induced by bootstrap sampling. Then, G_n^* consistently estimates G_n if and only if $T_n \xrightarrow{d} N(0, 1)$.

To see that the restriction to linear test statistics is not that restrictive, we illustrate an application of Mammen's theorem via an $M-$estimator (see e.g. Van der Vaart, 1998). θ is called an $M-$estimator if it is a solution to

$$\max_\theta \frac{1}{n} \sum_{i=1} q(X_i, \theta) = \max_\theta Q_n(\theta)$$

FOC :

$$
\begin{aligned}
\frac{\partial}{\partial\theta}Q_n(\theta_n) &= 0\\
\frac{\partial}{\partial\theta}Q_n(\theta_n) &= \frac{\partial}{\partial\theta}Q_n(\theta_0) + \frac{\partial}{\partial\theta\partial\theta'}Q_n\left(\tilde{\theta}\right)(\theta_n - \theta_0)
\end{aligned}
\tag{15}
$$

with $\tilde{\theta}$ between θ_0 and θ_n =)

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}\frac{\partial q(X_i;\theta_0)}{\partial\theta} &= \frac{1}{n}\sum_{i=1}\frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'}(\theta_n - \theta_0)\\
-\frac{1}{\sqrt{n}}\sum_{i=1}\frac{\partial q(X_i;\theta_0)}{\partial\theta} &= \frac{1}{n}\sum_{i=1}\frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n - \theta_0)
\end{aligned}
\tag{16}
$$

Since

$$\frac{1}{n}\sum_{i=1}\frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'} \to A \text{ (non-stochastic) a.s.} \tag{17}$$

we get

$$\sqrt{n}(\theta_n - \theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1} A^{-1}\frac{\partial q(X_i;\theta_0)}{\partial\theta} + o\left(n^{-1/2}\right) \tag{18}$$

We see that an $M-$estimators is asymptotically linear and therefore Mammen's theorem can be applied. Special cases of $M-$estimators are OLS ($q(x) = x^2$); Generalized Method of Moments (GMM), and Maximum Likelihood (ML), and Mammen's theorem can be applied to all these estimators.

Though the bootstrap is consistent in many often used situations, it can be inconsistent. A few examples of bootstrap inconsistency are

- Heavy tailed distributions

  Suppose X has a Cauchy distribution

  Define $\frac{1}{n}\sum_{i=1} X_i \xrightarrow{d} Z$, which is Cauchy distributed. The Cauchy distribution has no mean, therefore, we cannot use the Central Limit Theorem (CLT) or even the law of large numbers.

- Distribution of the sample average squared $\left(\overline{X}\right)^2$

  $\overline{X} = \frac{1}{n}\sum_{i=1} X_i;\qquad \mathbb{E}X_i = \mu$

  Now, distinguish two cases

– $\mu \neq 0$:

$$
\begin{aligned}
\left(\overline{X}\right)^2 &= \left[\frac{1}{n}\sum_{i=1}(X_i - \mu) + \mu\right]^2 \\
&= \mu^2\left[1 + \frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right]^2 \\
&= \mu^2\left[1 + \frac{2}{n\mu}\sum_{i=1}(X_i - \mu) + \left(\frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{n}\left(\left(\overline{X}\right)^2 - \mu^2\right) &= \frac{2}{\mu\sqrt{n}}\sum_{i=1}(X_i - \mu) + \sqrt{n}\left(\frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right)^2 \\
&\xrightarrow{d} N(0;1)
\end{aligned}
$$

– $\mu = 0$:

$$
\begin{aligned}
\mathcal{L}\left(n\left(\frac{1}{n}\sum_{i=1}X_i\right)^2\right) &= \mathcal{L}\left(\frac{1}{n}\sum_{i=1}X_i\right)^2 \\
&= \frac{1}{n}\left[N\left(0;\sqrt{n}\right)\right]^2 \\
&= \chi_1^2
\end{aligned}
$$

The second case violates the 3rd condition of Beran and Ducharme's (1991) theorem. This does not necessarily imply that the bootstrap is inconsistent at $\mu = 0$; but it can be shown that the bootstrap is inconsistent in this particular case.

Other examples are the estimation of a parameter on the boundary of the parameter space, the estimation of the maximum, and an AR(1) model with a unit root.

What can we do when the bootstrap is inconsistent? We can use the subsampling procedure introduced by Politis and Romano (1994a). Let the sample size be n. Now, let $m < n$. The statistic $T_n(X_1;\ldots;X_n) = \rho(n)\cdot[t_n(X_1;\ldots;X_n) - \theta]$. E.g. if $T_n = \sqrt{n}(\overline{X} - \theta)$ then $\rho(n) = \sqrt{n}$, $t_n = \frac{1}{n}\sum_{i=1} X_i$; and $\theta = \mathbb{E}(X_i)$. Suppose T_n has a non-degenerate asymptotic distribution. Consider subsets of the data, consisting of m observations. There are $\binom{n}{m} = N_m$ different subsets. Define the subset estimator for $G_n(\tau;F)$:

$$
G_{nm}(\tau) = \frac{1}{N_m}\sum_{k=1}^{N_m} 1\left[\rho(n)\cdot[t_{nm} - t_n] \leq \tau\right]; \qquad (19)
$$

where t_{nm} denotes the version of t_n for the subset.

Now, from Politis and Romano (1994a) it follows: If T_n has a nondegenerate asymptotic distribution, then G_{nm} estimates G_∞ consistently if $m \to \infty$, $n \to \infty$; and $\frac{m}{n} \to 0$.

If subsampling is always consistent, why use the bootstrap at all? When the bootstrap is consistent, the bootstrap estimator of $G_\infty(t;F)$ and $G_n(t;F)$ is (much) more accurate than the subsampling estimator. So, it is not wise to use subsampling when the regular bootstrap estimator is consistent.

4.2 Accuracy of the bootstrap

In the previous section, we defined the concept of consistency and some described conditions under which the bootstrap yields a consistent estimator of the distribution of the distribution of a statistic. This means, roughly speaking, that the bootstrap gets the asymptotic distribution right, at least if the sample size is large enough. Therefore we can use the bootstrap when we know that it is consistent and thereby avoid the often complicated mathematics of asymptotic theory. The bootstrap, however, often does more than just get the asymptotic distribution right. In a large number of relevant cases in econometrics, it provides us with a higher-order asymptotic approximation to the distribution of a statistic.

4.2.1 Bias reduction

We start by showing how the bootstrap can provide a reduction in the bias of an estimator. Consider a random variable X, with $\mathbb{E}(X) = \mu$. We want to estimate $\theta_0 = g(\mu)$, with g a known function (and μ an unknown parameter). We estimate μ by $\overline{X} = \frac{1}{n}\sum_{i=1} X_i$ and θ_0 by $\theta_n = g(\overline{X})$; which is sometimes called an analogue estimator. By Jensen's inequality, we know that θ_n will be biased for θ if g is a nonlinear function: $\mathbb{E}\theta_n = \mathbb{E}g(\overline{X}) \neq g(\mathbb{E}\overline{X}) = g(\mu) = \theta_0$. Therefore, θ_n is a biased estimator of θ_0.

Some people only focus on minimizing the bias of an estimator. However, from a decision theory perspective, the aim is to minimize the mean squared error of an estimator. As we know, the mean squared error is the sum of the bias squared and the variance of the estimator. Reducing the bias with increasing the variance may therefore not be beneficial for the mean squared error of the estimator. We will use the bias reduction argument here not for the sake of reducing the bias, but it is one of the two means to reduce the mean squared error. To see how the bootstrap can reduce the bias of θ_n assume that g is four times continuously differentiable and that the X has finite fourth absolute moments. A Taylor expansion of $\theta_n = g(\overline{X})$ gives

$$\begin{aligned} \theta_n - \theta_0 &= g(\overline{X}) - g(\mu) \\ &= g'(\mu)(\overline{X} - \mu) + \frac{1}{2} g''(\mu)(\overline{X} - \mu)^2 + R_n; \end{aligned} \tag{20}$$

where R_n is a remainder term that satisfies $\mathbb{E}(R_n) = o(n^{-2})$. Taking expectations on both sides of (20) results in

$$\mathbb{E}(\theta_n - \theta_0) = \frac{1}{2} \mathbb{E} g''(\mu)(\overline{X} - \mu)^2 + o\left(n^{-2}\right) \tag{21}$$

The first term on the right hand side of (21) has size $o(n^{-1})$. Therefore, through $o(n^{-1})$ the bias B_n of θ_n is

$$B_n = \frac{1}{2} \mathbb{E} g''(\mu)(\overline{X} - \mu)^2 \tag{22}$$

Now let us consider the bootstrap. With the bootstrap we sample the EDF of the data. Let $\{X_1^*, X_2^*, \ldots, X_n^*\}$ be the bootstrap sample that is obtained in this way. Define $\overline{X}^* = \frac{1}{n} \sum_{i=1} X_i^*$. The bootstrap estimator of θ is $g(\overline{X}^*) = \theta_n^*$, while the bootstrap "truth" is $g(\overline{X}) = \theta_n$. The bootstrap analog of (20) is

$$\begin{aligned} \theta_n^* - \theta_n &= g(\overline{X}) - g(\mu) \\ &= g'(\overline{X})(\overline{X}^* - \overline{X}) + \frac{1}{2} g''(\overline{X})(\overline{X}^* - \overline{X})^2 + R_n^*; \end{aligned} \tag{23}$$

where R_n^* is the bootstrap remainder term. Let $\mathbb{E}^*$ denote the expectation with respect to the EDF of the estimation data. Taking $\mathbb{E}^*$ expectations of (23) gives

$$B_n^* = \frac{1}{2} \mathbb{E}^* g''(\mu)(\overline{X}^* - \overline{X})^2 + o\left(n^{-2}\right) \text{ a.s.} \tag{24}$$

By a Monte Carlo simulation we can compute B_n^* with arbitrary accuracy, since the distribution of the bootstrap samples is known. This Monte Carlo procedure is as follows

1. Use estimation data to compute θ_n.

2. Generate a bootstrap sample of size n by sampling data randomly with replacement. Compute $\theta_n^* = g\left(\overline{X}^*\right)$.

3. Compute $\mathbb{E}^*(\theta_n^*)$ by averaging results of many repetitions of step 2. Set $B_n^* = \mathbb{E}^*(\theta_n^*) - \theta_n$.

We have

$$\mathbb{E}(B_n^*) = B_n + O\left(n^{-2}\right)$$

This implies that if we use a modified estimator of θ_0; $\tilde{\theta}_n = \theta_n - B_n^*$ we can get a bias reduction from $O(n^{-1})$ to $O(n^{-2})$, since

$$\begin{aligned}\mathbb{E}\left(\tilde{\theta}_n - \theta\right) &= \mathbb{E}(\theta_n - \theta) - \mathbb{E}(B_n^*)\\ &= B_n - B_n + O\left(n^{-2}\right)\\ &= O\left(n^{-2}\right).\end{aligned}$$

Note that, we can calculate B_n^* as accurate since we control the number of repetitions of the Monte Carlo experiment. We illustrate the theory with an example in Section 6.1.

4.2.2 Distribution of statistics

Though we have shown that the bootstrap is consistent, we still haven't discussed the difference between $G_n(t; F_n)$ and $G_n(t; F_0)$. This can be done using Edgeworth expansions. For a rigorous treatment, this involves some tedious mathematics that will not be presented here. Instead we only provide a small portion of the mathematics that should provide the necessary insights. Consider

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1} X_i;\ \mathbb{E}X_i = 0;\ \mathrm{Var}X_i = 1 \tag{25}$$

Note that the restrictions on the mean and variance are mild, since by standardization we can transform most statistics in this way. Using the CLT, we know that $\lim_{n \to \infty} \mathbb{P}(S_n \le z) = \Phi(z)$; where $\Phi(\cdot)$ is the cumulative normal distribution function. What can we say about the difference $\mathbb{P}(S_n \le z) - \Phi(z)$? Under some regularity conditions on G_n we find that

$$\mathbb{P}(S_n \le z) = \Phi(z) + \frac{1}{\sqrt{n}} g_1(z) + \frac{1}{n} g_2(z) + \ldots + n^{-j/2} g_j(z) + O\left(n^{-(j+1)/2}\right) \tag{26}$$

where $g_j(\cdot)$ is an even function of t of j is odd and an odd function of t when j is even[7]. After some calculations using Fourier inversions, we get

[7] A function is called even if $f(x) = f(-x)$ and is called odd if $f(x) = -f(-x)$.

$$p_n(z) = \phi(z)\left(1 + \frac{1}{6}\mathbb{E}\left(X^3\right)\frac{z^3 - 3z}{\sqrt{n}}\right). \tag{27}$$

For $n \to \infty$ we get the standard normal density function, since the second term between brackets goes to zero. For finite samples, however, we find an improvement taking third moments into account.

Assume the statistic of interest is a smooth function of the sample moments

$$T_n = \frac{\sqrt{n}(H(\overline{X}) - H(\mu))}{s} \tag{28}$$

where s^2 is a consistent estimator of the asymptotic variance of $\sqrt{n}(H(\overline{X}) - H(\mu))$. Generally, this will be unknown, if it is known (e.g. σ^2); we use σ^2 instead. Some examples of functions H that cannot be used are the least absolute deviation (not sufficiently smooth), and nonparametric kernel density and regression estimators (depends on sample size). Under some regularity conditions, we find

$$G_n(t;F) = G_\infty(t;F) + \frac{1}{\sqrt{n}}g_1(t;F) + \frac{1}{n}g_2(t;F) + \frac{1}{n\sqrt{n}}g_3(t;F) + o\left(n^{-2}\right) \tag{29}$$

where $g_j(t;F)$ is an even function of t if j is odd and an odd function if j is even. Also, g_j is a continuous functional of its second argument F_0. We can do the same expansion for the bootstrap estimator

$$G_n(t;F_n) = G_\infty(t;F_n) + \frac{1}{\sqrt{n}}g_1(t;F_n) + \frac{1}{n}g_2(t;F_n) + \frac{1}{n\sqrt{n}}g_3(t;F_n) + o\left(n^{-2}\right) \tag{30}$$

The error made by the bootstrap estimator is then

$$\begin{aligned} G_n(t;F_n) - G_n(t;F) &= G_\infty(t;F) - G_\infty(t;F_n) \\ &\quad + \frac{1}{\sqrt{n}}(g_1(t;F_n) - g_1(t;F)) + \frac{1}{n}(g_2(t;F_n) - g_2(t;F)) + o\left(n^{-3/2}\right). \end{aligned} \tag{31}$$

The leading term of the error made by the bootstrap is $G_\infty(t;F) - G_\infty(t;F_n)$. We know that $\rho(F_n;F) = o_p(n^{-\frac{1}{2}})$, therefore the bootstrap estimator and the asymptotic approximation make an equally large error. If the test statistic is, however, asymptotically pivotal, i.e. $G_\infty(t;F) = G_\infty(t)$, the leading term of (31) equals zero. Then the new leading term of the bootstrap error becomes $\frac{1}{\sqrt{n}}(g_1(t;F_n) - g_1(t;F))$ which is $o_p(n^{-1})$. Thus, in the case of asymptotically pivotal statistics the bootstrap is more accurate than a first-order asymptotic approximation. We call this an asymptotic refinement. It can be shown that the bootstrap becomes even better if the distribution is symmetric, since then the functions g_2 and g_4 drop out. The leading term is then of order $o\left(n^{-3/2}\right)$. We may therefore conclude that the bootstrap is more accurate that first order asymptotic approximation (e.g. asymptotic normal distribution) for estimating the distribution of an asymptotically pivotal statistic.

How do we implement the bootstrap for computing critical values? We have as statistic

$$T_n = \sqrt{n}\,(\theta_n - \theta_0)/s_n$$

where θ_n is the parameter estimator, θ_0 hypothetical value ($H_0 : \theta = \theta_0$), and s_n the standard error of asymptotic distribution of $\sqrt{n}\,(\theta_n - \theta_0)$.

The Monte Carlo procedure for computing bootstrap critical values.

1. Use estimation data to compute θ_n.

2. Generate a bootstrap sample of size n by sampling the distribution corresponding to F_n. (E.g. F_n is the EDF of the data, draw the data randomly with replacement.) Compute estimators θ_n^* and s_n^* from the bootstrap sample. Form the bootstrap test statistic $T_n^* = \sqrt{n}\,(\theta_n^* - \theta_n)/s_n^*$.

   Note: We do not center the bootstrap statistic around θ_0. This does not converge to zero, since the null hypothesis concerns the population, and not the data. (There is a zero probability of obtaining $\theta_n = \theta_0$, but we like to test whether they are not "too far apart".)

3. Use the results of many repetitions of step 2 to compute the empirical distribution of $|T_n^*|$ (assuming a symmetrical test here). Set z_α^* equal to the $(1-\alpha)$-quantile of this distribution.

Example: Construct a nominal 95% confidence interval for $\theta = \exp\{\mu\}$, when $X \sim N(0;6)$. The true value $\theta_0 = 1$. Now, $\theta_n = \exp\{\bar{x}\}$, and we take $n = 10$. The procedure is as follows:

1. Generate estimation data set of size $n = 10$ by sampling $N(0;6)$.

2. Compute z_α^* by following steps 1–3 from the previous procedure.

3. Find the empirical coverage probabilities of the bootstrap and asymptotic confidence intervals by repeating the steps 1 and 2 of this procedure many times (e.g. 1000).

In Horowitz (1999), Horowitz finds with a nominal coverage probability of 95%, 94.3% for the bootstrap and 88.5% for the asymptotic approximation. In Section 6.2 we perform a simulation study for the classical regression model.

5 Selected topics

In this section we discuss some miscellaneous topics concerning the bootstrap. First, we discuss the topic of recentering in the context of GMM. Second, we look at the bootstrap in case of time series. Finally, we discuss some cases where the bootstrap flops.

5.1 Recentering

An often used econometric technique these days is the Generalized Method of Moments (GMM) estimation introduced by Hansen (1982). This happens to be a technique where the bootstrap does not necessarily provide an asymptotic refinement. To see why, let θ_0 be the true value of a parameter θ that is identified by the moment conditions $\mathbf{E}h(X;\theta) = 0$, where h is a known function. The dimension of h is $k \times 1$; and the dimension of θ is $q \times 1$; where $k \geq q$: We might want to estimate θ by solving $\frac{1}{n}\sum_{i=1} h(X_i;\theta) = 0$: In the case that $k = q$, this means that θ is exactly identified. There is no exact solution to these set of restriction when $k > q$; therefore we need to solve

$$\min_{\theta} \sum_{i=1} h(X_i;\theta)' \Omega_n \sum_{i=1} h(X_i;\theta): \tag{32}$$

where Ω_n is a weighting matrix. The solution to this problem is the GMM estimator, say θ_n: A natural way to proceed, would be to use the following bootstrap version of (32)

$$\min_{\theta} \sum_{i=1} h(X_i^*;\theta)' \Omega_n \sum_{i=1} h(X_i^*;\theta): \tag{33}$$

This would, however, be naive. The true value in the bootstrap world is namely, θ_n, not θ_0: This implicitly assumes that the moment conditions are exactly zero, but in the case of overidentifying restrictions the moment restrictions are not exactly zero.

$$\frac{1}{n}\sum_{i=1} h(X_i;\theta) \neq 0: \tag{34}$$

The problem can be resolved by recentering. Introduce

$$h^*(X_i;\theta) = h(X_i;\theta) - \frac{1}{n}\sum_{i=1} h(X_i;\theta) \tag{35}$$

We now have a new set of bootstrap moment restrictions

$$\mathbb{E}h^*(X_i;\theta) = 0. \tag{36}$$

This is the set of moment conditions we should use to get a bootstrap estimator θ_n^*. An insightful example is the classical regression without intercept.

Consider a linear model with fixed design (X deterministic) and no intercept, i.e.

$$Y_i = \beta X_i + \varepsilon_i$$

and $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$. Now, estimate β by OLS to get estimator b and residuals $e_i = Y_i - bX_i$. We would like to use the bootstrap model $Y_i^* = \beta X_i^* + \varepsilon_i^*$ with $\mathbb{E}(\varepsilon_i^*) = 0$ and $\mathrm{Var}(\varepsilon_i^*) = \sigma^{*2}$. Since there is no constant term, the residuals do not necessarily sum to zero. Therefore, in the bootstrap model $\mathbb{E}(\varepsilon_i^*) \neq 0$. If we recenter the residuals, i.e. $\tilde{\varepsilon}_i^* = \varepsilon_i^* - \frac{1}{n}\sum_{i=1} e_i$ we get the model $Y_i^* = \beta X_i^* + \tilde{\varepsilon}_i^*$ with $\mathbb{E}(\tilde{\varepsilon}_i^*) = 0$ and $\mathrm{Var}(\tilde{\varepsilon}_i^*) = \sigma^{*2}$ which allows us to use the bootstrap.

5.2 Time series data

So far, we considered cases in which the data were i.i.d. Here we investigate the case where the data are dependent. Bootstrap sampling must then be done in a way that captures the dependence structure of the data. We consider two cases:

- If there is a parametric model available that reduces the DGP to independent random sampling, then i.i.d. results continue to hold. Consider, for example, an ARMA model

$$A(L;\theta_0)Y_t = B(L;\psi_0)\varepsilon_t \tag{37}$$

where L denotes the lag operator and θ_0, ψ_0 parameters. We estimate θ_0 by θ_n, ψ_0 by ψ_n and obtain the residuals $e_t = (B(L;\psi_n))^{-1}(A(L;\theta_n)Y_t$. As we saw in the previous section, we want residuals that are zero on average in the bootstrap world, therefore $e_t^* = e_i - \frac{1}{n}\sum_{i=1} e_i$ can be used. Now we can generate bootstrap series by simulating $Y_t^* = ((A(L;\theta_n))^{-1}B(L;\psi_n)e_t^*$.

- If there is no parametric model available, sampling in blocks is an alternative. This can be done by either non-overlapping blocks or overlapping blocks. Consider, for example, a time series with T observations $\{Y_1; \ldots; Y_T\}$. We can take blocks of length l, in the case of non-overlapping blocks block 1 is $\{Y_j; Y_{j+1}\}$; block 2 is $\{Y_{j+2}; Y_{j+3}\}$; etc. With overlapping blocks block 1 is $\{Y_j; Y_{j+1}\}$; block 2 is $\{Y_{j+1}; Y_{j+2}\}$; etc. The bootstrap sample is obtained by drawing blocks with replacement and laying them end-to-end in the order sampled. A more advanced method is the stationary bootstrap by Politis and Romano (1994b). In this case we use overlapping blocks with lengths drawn from the geometric distribution.

  Though the block bootstrap provides some asymptotic refinements, it is less attractive that in the i.i.d. case.

  – With the block bootstrap and a symmetrical confidence interval test we obtain $\mathbb{P}\left(|T_n| > z^*_{n\alpha/2}\right) = \alpha + o\left(n^{-\frac{4}{3}}\right)$. The bias is thus of order $o\left(n^{-\frac{4}{3}}\right)$. If we compare this with the i.i.d. case where we had $o(n^{-2})$; we see that we lose quite some efficiency gain, since the asymptotic case already yields $o(n^{-1})$.

  – We must use special forms of the test statistics, we need correction factors to account for the distortion caused by blocking. The reason for this is that time series relations are deterministic when blocks are formed.

5.3 The bootstrap can be a flop

Though the bootstrap often provides us with improved results, it should not be used blindly. As we saw, the bootstrap only leads to asymptotic refinements if a test statistic is asymptotically pivotal, if not, first-order asymptotic distribution theory approximations can be used just as well. It must be said that the bootstrap is still useful in this case if the first order asymptotic distribution theory approximation is difficult to compute. There are, however, also situation when the bootstrap should not be used:

1. If it is inconsistent, we saw examples of this in Section 4.1.

2. Instrumental variables estimation with weak instruments. Consider the model, $Y = X\beta + \varepsilon$; instrument Z; this results in IV estimator

$$b_n^{IV} = \frac{Z'Y}{Z'X} \tag{38}$$

If we have weak instruments, $\mathbb{E}(Z'X) \approx 0$. Therefore, in small samples, $Z'X$ may bounce around zero and consequently b_n^{IV} may be a very large positive or negative number.

6 Applications

In this section, we implemented some of the bootstrap theory presented above. First, we apply the bootstrap for bias reduction of an estimator as discussed in Section 4.2.1. Second, we look at the classical regression model we discussed in Section 2[8].

6.1 Bias reduction

In Section 4.2.1 we discussed the issue of bias reduction. Here we present an example to illustrate this. The presented example is an extended version of Example 3.2 in Horowitz (1999). We have $X \sim N(0,6)$, $n = 10$, $\theta = g(\mu) = \exp\{\mu\}$, $\theta_0 = 1$, and $\theta_n = \exp\left(\overline{X}\right)$. We perform the following procedure:

1. Generate estimation data set of size $n = 10$ by sampling the normal distribution and use this data to compute θ_n.
2. Compute B_n^* and form $\tilde{\theta}_n = \theta_n - B_n^*$.
3. Estimate $\mathbb{E}(\theta_n - \theta_0)$ and $\mathbb{E}(\theta_n - B_n^* - \theta_0)$ by averaging results of many replications of steps 1 and 2.

In Tables 1 and 2 we see that the bootstrap provides high improvements on the MSE when the sample size is small. If we increase the sample size to 100, the results are less shocking, but still the bootstrap outperforms the analogue estimator. In Table 2 we see that increasing the number of bootstrap replications from 1000 to 5000 does not really improve our results. We have also checked the robustness of the bootstrap method for other values of σ^2, and found that the bootstrap provides similar improvements here. One is encouraged to perform these experiments for oneself using the program provided.

[8] Ox programs to perform the experiments in this section can be found on the author's web page: http://cwis.kub.nl/~few5/center/phd_stud/kerkhof/

	Horowitz			Gauss			Ox		
	Bias	MSE	Var	Bias	MSE	Var	Bias	MSE	Var
θ_n	0.356	1.994	1.867	0.274	1.149	1.074	0.3280	1.5670	1.6746
$\theta_n - B_n^*$	-0.063	1.246	1.242	-0.126	0.690	0.674	-0.0986	0.8424	0.8521

Table 1: Simulation results for the estimation of $\theta = \exp(\mu)$, with $\theta_n = \exp\left(\bar{X}\right)$. Furthermore, $X \sim N(0, 6)$, and $n = 10$. The first column restates the Table from Horowitz (1999). In the second column, the results from the Gauss program by Horowitz are displayed. The third column shows the results from an Ox program written by the author. In the first row, the analogue estimator is displayed, and in the second row the bootstrap estimator. Number of simulations used is 10,000 and the number of bootstrap replications equals 5,000.

			bootstrap = 5000		bootstrap = 1000	
n	σ^2	rep	Bias	MSE	Bias	MSE
10	6	$10,000$	0.3280	1.6746	0.3206	1.5060
			-0.0986	0.8521	-0.0994	0.7717
100	6	$10,000$	0.0256	0.0657	0.0254	0.0661
			-0.0053	0.0613	-0.0056	0.0617

Table 2: Results to check the influence of the sample size on the results.

6.2 Confidence intervals

We now return to the example of classical regression that we gave at the start of this paper. We want to test whether certain parameter estimates differ significantly from zero. Usually this is done using a (1-α)% confidence interval and checking whether 0 is in this interval. We now set up a Monte Carlo experiment to check the performance of the bootstrap for the classical regression example. Recall the model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{39}$$

with an i.i.d. sequence $\{\varepsilon_i\}$ such that

$$\begin{aligned} \mathbb{E}(\varepsilon_i) &= 0 \\ \mathbb{E}(X_i \varepsilon_i) &= 0 \\ \mathbb{E}\left(\varepsilon_i^2\right) &= \sigma^2. \end{aligned} \tag{40}$$

We simulated random samples of this model with $\beta_0 = \beta_1 = 0$; and 1) $\mathcal{L}(\varepsilon_i) = N(0,1)$ and 2) $\mathcal{L}(\varepsilon_i) = U[-1,1]$. In both cases $\mathcal{L}(X_i) = N(0,1)$. We construct 4 confidence intervals based on

1. Asymptotically critical value with normal standard errors.
2. Asymptotically critical value with bootstrap standard errors.
3. Bootstrap critical value with normal standard errors.
4. Bootstrap critical value with bootstrap standard errors.

In the case where $\mathcal{L}(\varepsilon_i) = N(0,1)$, we know the exact confidence interval, since the critical value is known (from the t $-$ distribution). The procedure is as follows:

a. Generate a random sample $\{(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$.

b. Construct L bootstrap samples: $\{\{(Y_1^*, X_1^*), (Y_2^*, X_2^*), \ldots, (Y_n^*, X_n^*)\}_j, j = 1, \ldots, L\}$ by drawing pairs (Y_i, X_i) with replacement from the data.

c. Calculate the OLS estimators $\left\{b_j^*, j = 1, \ldots, L\right\}$ for the bootstrap samples.

d. Compute bootstrap standard errors $\left\{\sigma_j^* = \sqrt{\operatorname{diag}\left(\operatorname{Var}\left(b_j^*\right)\right)}, j = 1, \ldots, L\right\}$.

e. Compute bootstrap critical values $\left\{z_{n\alpha,j}^*, j = 1, \ldots, L\right\}$.

	A	B	A^*	B^*
intercept	0.901	0.960	0.917	0.964
slope	0.903	0.940	0.937	0.951

Table 3: Coverage ratios based on 4 types of confidence intervals for the case with $L(\varepsilon_i) = N(0;1)$; A : asymptotic critical value with normal standard errors. A^* : asymptotic critical value with bootstrapped critical value. B : bootstrap critical value with normal standard errors. B^* : bootstrap critical value with bootstrapped critical value.

	A	B	A^*	B^*
intercept	0.896	0.959	0.909	0.962
slope	0.901	0.928	0.930	0.940

Table 4: Coverage ratios based on 4 types of confidence intervals for the case with $L(\varepsilon_i) = U[-1;1]$; A : asymptotic critical value with normal standard errors. A^* : asymptotic critical value with bootstrapped critical value. B : bootstrap critical value with normal standard errors. B^* : bootstrap critical value with bootstrapped critical value.

f. Construct confidence intervals and check whether the original parameter estimates are within the confidence interval:

We see in Tables that based on the asymptotic critical value (both using normal standard errors as using bootstrapped standard errors), we reject a null hypothesis too often. The coverage ratio in the case of bootstrap critical values is much closer to the wanted 95%. The results are approximately the same for the both case 1) $L(\varepsilon_i) = N(0;1)$ and 2) $L(\varepsilon_i) = U[-1;1]$. We may therefore conclude that as indicated by the theory, the bootstrap provides us with a better coverage ratio if the we base our decision on the confidence interval using bootstrap critical values. We also see that bootstrap standard errors do not provide us with a better coverage ratio, though it doesn't underperform either.

7 Conclusions

In this paper, we suggest the bootstrap as an alternative to the commonly used first-order asymptotic theory. Though the idea of the bootstrap has been around since Efron (1979), it only became popular the last couple of years. This is not surprising, since for a practical implementation the bootstrap relies on heavy computer power.

In this paper, we started by motivating the use of the bootstrap in probably the most used econometric model, the classical regression. We showed in Section 6.2 that the bootstrap improves upon the asymptotic con¯dence interval for the t-statistics. We continued to treat the general theory behind the bootstrap and its main idea "treating the data as if it were the population". We discussed its properties and illustrated these with examples. To show that the bootstrap also has its de¯ciencies, some examples of bootstrap inconsistencies were provided.

Finally, we performed some simulation studies to investigate the performance of the bootstrap versus that of asymptotic theory. We showed that the bootstrap can provide us with bias reductions of estimators and asymptotic re¯nements of distributions of statistics.Considering the theoretical advantages combined with increasing computer power we can only expect the bootstrap to become a standard tool in econometrics.

References

Beran, R., and G. Ducharme (1991): Asymptotic Theory for Bootstrap Methods in Statistics. Centre de recherches mathematique, Universite de Montreal, Montreal, Canada.

Efron, B. (1979): \Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7, 1{26.

Hansen, L. P. (1982): \Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50, 1029{1054.

Horowitz, J. L. (1999): \The Bootstrap," Handbook of Econometrics, vol. 5, pp. 1{98.

Mammen, E. (1992): When Does the Bootstrap Work? Springer-Verlag, New York.

Politis, D., and J. Romano (1994a): \Large Sample Con¯dence Regions Based on Subsamples under Minimal Assumptions," Annals of Statistics, 22, 2031{2050.

Politis, D., and J. Romano (1994b): \The Stationary Bootstrap," Journal of the American Statistical Association, 89, 1303{1313.

Financial Crises: Theories and Models
Hyun Song Shin

Report by Remco van der Molen, University of Groningen[¤]

1 Introduction

Financial crises are infamous for their often disastrous consequences. Moreover, there does not always seem to be a relationship between a ¯nancial crisis and changes in the economic fundamentals, which makes these crises hard to predict. To explain this feature, economists have pointed to the self-ful¯lling nature of ¯nancial crises. In economic theory, a ¯nancial crisis is often viewed as a random shift from one equilibrium to the other. However, this argument is only an informal one. What triggers these self-ful¯lling beliefs, and what should economists advise politicians? These questions cannot be addressed by theories which point only to the random nature of ¯nancial crises. In his lectures, professor Shin introduced a theory and some models that can give more insight in the onset of ¯nancial crises and their probability of occurrence. First, he freshened our knowledge of the more traditional models of ¯nancial crises (section 2) and evaluated them (section 3). In order to improve on these models, professor Shin used the theory of global games (section 4). The rather technical introduction in this game-theoretic concept proved to be worthwhile when professor Shin showed some very interesting applications (sections 5 and 6).

2 First and second generation models

After the break-down of the Bretton Woods system, numerous countries have tried to peg their exchange rate at some ¯xed level or to some other currency's exchange rate. Almost none of these attempts turned out a success; eventually, the peg was given up. In a lot

[¤]Department of Economics, University of Groningen, PO Box 800, 9700AV Groningen.

of cases, this happened after a speculative attack. In this period, several models evolved that tried to explain these stylized facts. Two important models are discussed below.

2.1 The Krugman (1979)-model

In this model, the central bank has two di®erent tasks. Firstly, it has to defend the exchange rate by holding the monetary base ¯xed. Its second objective is to monetize (part of) the exogenously given ¯scal de¯cit by buying domestic currency bonds b. De¯ne the change in domestic currency holdings by the central bank $\dot{b}_t$ = ¹. Now, in order to keep the monetary base ¯xed, the central bank has to soak up its foreign currency reserves. Obviously, this policy is not sustainable, for at time T, say, the bank will run out of foreign currency reserves. From that time on, the central bank will no longer be able to keep the exchange rate at the ¯xed level. So the interesting question in the model is not whether the peg will fail, but rather when this will occur.

Once the central bank runs out of foreign reserves, at time T, the money supply will begin to rise at rate ¹. This means that the expected rate of depreciation will jump from zero to ¹ instantaneously, causing the demand for real money balances to drop sharply. The discrete change in the exchange rate at time T is of course anticipated by speculators, and so they will all switch out of domestic currency into foreign reserves before time T. This explains that in this setting a speculative attack is inevitable. It is obvious, however, that this speculative attack will not occur at time T, precisely because it is anticipated. To see when it will occur, Krugman introduces the concept of the shadow exchange rate, which is the °oating exchange rate that would prevail if all the central bank's foreign assets had already passed into private hands. In other words, the speculators, who know that the central bank's policy is not sustainable, have already discounted this knowledge in their shadow exchange rate. In our model this shadow exchange rate rises over time at rate ¹. So there is one point where the shadow exchange rate is exactly as high as the ¯xed exchange rate. Precisely at this point the speculative attack will occur, because this is the only point where the change from a ¯xed to a °oating exchange rate can occur without a discontinuous change. Therefore, this is the only point that is consistent with perfect foresight.

Note that in the Krugman model, perfect foresight is assumed in case of the speculators, whereas the central bank does not seem to care about the long-term sustainability of its policy. There is no a priori reason why the central bank would be myopic. Moreover, speculative attacks typically seem to have a self-ful¯lling nature, in that an exchange rate will collapse if enough people expect it to collapse. However, this self-ful¯lling nature is not captured by the Krugman model.

2.2 The Obstfeld (1996)-model

The trade-off between two conflicting objectives of the government that was central in the Krugman model is also central in the more general model by Obstfeld. In this model, the monetary authority may for example try to avoid a currency attack by raising the interest rates, but this may have adverse effects on the domestic economy. The pain of sticking to the fixed exchange rate will in general be greater the heavier the attack.

In the Obstfeld model, the conflict between the "domestic targets" and sticking to a fixed (but adjustable) exchange rate is made explicit in a loss function,

$$L = (y - \tilde{y})^2 + \chi\pi^2 + c(\pi) \tag{1}$$

where y is output, $\tilde{y}$ is the output target, χ is the relative weight the authorities place on inflation stabilization versus employment stabilization, $\pi = \dot{e}$ is the change in exchange rate or inflation and $c(\pi)$ is the cost of changing the currency peg. Furthermore, output is determined by an expectation augmented Phillips curve:

$$y = \bar{y} + (\pi - \pi^e) - z \tag{2}$$

where $\bar{y}$ is the natural level of output, π^e is the private sector's belief of π, and z is the noise term. So, deviation of output from its natural rate depends on unexpected changes in the exchange rate and on random shocks. The private sector chooses π^e before observing either z or π, whereas the government chooses π after observing both z and π^e.

The government defends the fixed exchange rate as long as the random shocks are not too large in absolute value. When the shocks are too large, using monetary policy for keeping the exchange rate fixed is too costly, and the government rather uses it to stabilize output. We can see that in this model the relationship between the rational expectation of inflation and the private sector's belief about inflation, π^e, may be very complicated. The problem is that a rise in inflation expectations raises the cost of keeping the exchange rate fixed. This makes it less unattractive for the government to lower the exchange rate indeed, and therefore raises the probability of adjustment of the exchange rate. Since the equilibrium condition is that the rational expectation and the private sector's belief about inflation coincide, there may very well be multiple equilibria in this model. And because of the existence of multiple equilibria, the government may not be able to enforce its most favored equilibrium. A small change in expectations may already cause a shift from one equilibrium to another. Such a jump would look very much like a speculative attack in the Krugman model, forcing the government to abandon the currency peg.

3 The third generation?

The most important contribution of the above models is that they have very clearly pointed out that defending a currency peg entails costs. Therefore, defending cannot go on forever. Whatever the perceived bene¯ts and whatever the public pronouncements, there is a pain threshold at which the costs of defending the peg outweigh the bene¯ts.

Furthermore, the Obstfeld model does shed some light on the self-ful¯lling nature of currency attacks. It also leaves open some important issues, however. A jump from one equilibrium to another can only be explained in an informal manner; a more formal explanation would be preferable. Moreover, the model assumes that the private sector can be represented by a single agent. The assumption does no longer hold if there are pay-o® interactions between the agents, because then not only the agent's own beliefs matter, but also the beliefs of others. This situation seems to be typical for currency attacks, where the optimal action depends on what other agents choose.

This can easily be seen by looking at the many parties that play a role in currency attacks, e.g. speculators, domestic ¯rms, domestic banks, depositors, foreign banks. If one of the parties undertakes a painful action, i.e. an action that makes it more costly to keep the exchange rate ¯xed, it is in the interest of the other parties that they also choose painful actions. Take, for example, the case where a government has pegged its currency to the US dollar. When speculators have information that makes them believe that the domestic currency will devalue, they will borrow the domestic currency and buy dollars. In response to this, domestic ¯rms and banks will also start to sell the domestic currency, because of hedging purposes. Hence, actions which increase the pain are mutually reinforcing. So a realistic model of currency attacks would have to take into account that the beliefs of other players matter.

Another feature of the Obstfeld model is that an attack on a government with weak fundamentals (low costs of adjusting the exchange rate, a high output target, low Â) is more likely to lead to a shift in equilibria. So, given a speculative attack, a government with strong in°ation fundamentals is more likely to resist this attack than one with weak fundamentals. The model does not tell us anything, however, about the relationship between the fundamentals and the private sector's beliefs. In this model, the beliefs can be chosen independently of the fundamentals, while it seems more plausible to assume that the underlying fundamentals have some in°uence on the private sector's belief about in°ation.

4 Global Games

Central in the Morris and Shin (1998) model of currency attacks is the concept of global games, first introduced by Carlsson and Van Damme (1993). It is this theory that enables them to derive a unique equilibrium. Multiple equilibria outcomes often depend crucially on the assumptions of common knowledge of the fundamentals and perfect knowledge of the other agents' actions in equilibrium. But one may argue that common knowledge is too simple a picture. Although market participants do seem to be very well informed because of the availability of a lot of information, this information may still be imperfect. Moreover, because different market participants may use different sources of information, there may be disparities in their information. Therefore, even when everyone knows the fundamentals, they may not know that everyone knows the fundamentals. Still higher orders of uncertainty may be relevant. It is thus unlikely that common knowledge adequately describes the underlying information structure. We will see that changing the assumptions of common knowledge and perfect information solves the multiple equilibrium problem, and results in a model that can address the issues that were unexplained by the first and second generation models.

4.1 Perfect information

Consider first the following perfect information benchmark model. There are two identical investors, each with an endowment of one unit. This unit can either be consumed or invested in a risky project. The output per head from the risky investment is

$$\begin{cases} R & \text{if both invest} \\ bR & \text{if only one invests} \end{cases}$$

where R is a lognormal random variable, and with $0 < b < 1$ and $\log b = -1$. We assume that R is chosen by nature at the beginning of the game, after which it becomes common knowledge, and the players decide wether or not to invest. Note that investing is mutually complementary, in the sense that the return to investment is higher when both invest. Both investors have a logarithmic utility function $u(c) = \log c$, so the utility of consuming the endowment is 0. When we define $r = \log R$, we get the following payoff matrix:

	Invest	Refrain
Invest	$r; r$	$r - 1; 0$
Refrain	$0; r - 1$	$0; 0$

Observe that for $0 < r < 1$ there are two pure equilibria, (Invest, Invest) and (Refrain, Refrain). When we extend the game to more players and mixed equilibria, there can be very many equilibria. So, ex ante, the outcome of this game is indeterminate. However, the Pareto efficient outcome is that both players invest whenever $r \geq 0$.

4.2 Imperfect information

Now we change the information structure of the game. The return on investment, R, is no longer common knowledge after it is chosen by nature. Rather, there is uncertainty about the returns. Instead of knowing the return exactly, the players receive noisy signals about R. Based on these signals, the players decide whether to invest or to refrain. r is normally distributed with mean $\bar{r}$ and variance $1/\alpha$: $r \sim N(\bar{r}; 1/\alpha)$. $\bar{r}$ is assumed to be common knowledge. A larger α thus reflects less uncertainty about the return to investment. So r can be viewed as reflecting the state of nature. Investor i receives a noisy signal x_i of this state of nature, $x_i = r + \varepsilon_i$, with $\varepsilon_i \sim N(0; 1/\beta)$ and the ε_i are independent. When β gets large, the signal becomes more precise.

Based on the signal x_i, player i forms an estimate of r, ρ_i:

$$\rho_i = E(r \| x_i) = \frac{\alpha\bar{r} + \beta x_i}{\alpha + \beta} \tag{3}$$

This posterior belief about r is player i's best estimate of r. Given the prior mean of r, there is a one to one correspondence between ρ_i and x_i. The expected payoff for investor i conditional on x_i is $E(r - I|\rho_i)$, where I is an indicator function for the other player's investment, which has the following form:

$$I = \begin{cases} 1 & \text{if opponent refrains} \\ 0 & \text{if opponent invests} \end{cases}$$

From the definition of the posterior belief, it follows that the expected conditional payoff can be rewritten as

$$E(r - I|\rho_i) = \rho_i - E(I|\rho_i) \tag{4}$$

So player i's expected payoff depends on his posterior belief and on his expectation about the action of his opponent conditional on his posterior belief. A strategy for player i is a function $\rho_i \mapsto \{\text{Refrain, Invest}\}$. When we think of the signal x_i that player i receives as his type, we can solve this game for the Bayesian Nash equilibria. The equilibrium strategies have the form of a pair of strategies such that the investor's expected payoff is maximized, conditional on his posterior belief ρ_i, and given the strategy of the other investor.

4.2.1 Switching strategies

We will start solving this game by concentrating on switching strategies. So we want to find an equilibrium that consists of strategies of the following form

$$\begin{cases} \text{Invest if } \rho_i \geq \rho^* \\ \text{Refrain if } \rho_i < \rho^* \end{cases} \tag{5}$$

According to this strategy, player i should invest if his posterior belief about r is high enough, and refrain otherwise. In order for this strategy to be an equilibrium strategy, player i must be indifferent between investing and refraining when his posterior belief is at the switching point ρ^*. In other words, the expected conditional payoff of investing, $E(r - I|\rho^*)$, must equal that of refraining, 0. From (4), we know that $E(r - I|\rho^*) = \rho^* - E(I|\rho^*)$. The indicator function I takes the value 1 if and only if $\rho_j < \rho^*$, and therefore $E(I|\rho^*) = \Pr(\rho_j < \rho^*|\rho^*)$. Player j's posterior belief will be below the switching point if and only if

$$\frac{\alpha\bar{r} + \beta x_j}{\alpha + \beta} < \rho^* \tag{6}$$

We can rewrite this as $x_j < \rho^* + \frac{\alpha}{\beta}(\rho^* - \bar{r})$. Based on ρ^*, player i can infer player j's signal x_j. Conditional on ρ^*, r is normally distributed with mean ρ^* and variance $\frac{1}{\alpha+\beta}$ (because of the independence of the error terms). Therefore,

$$x_j|\rho^* \sim N\left(\rho^*; \frac{1}{\alpha+\beta} + \frac{1}{\beta}\right) = N\left(\rho^*; \frac{\alpha + 2\beta}{\beta(\alpha+\beta)}\right) \tag{7}$$

Now we can calculate the probability,

$$\begin{aligned} \Pr(\rho_j < \rho^*|\rho^*) &= \Pr(x_j < \rho^* + \tfrac{\alpha}{\beta}(\rho^* - \bar{r})|\rho^*) \\ &= \Phi\left(\sqrt{\frac{\beta(\alpha+\beta)}{\alpha+2\beta}}\left(\rho^* + \frac{\alpha}{\beta}(\rho^* - \bar{r}) - \rho^*\right)\right) \\ &= \Phi(\sqrt{\gamma}(\rho^* - \bar{r})) \end{aligned} \tag{8}$$

where Φ is the cumulative normal distribution centered on the ex ante mean $\bar{r}$, and

$$\gamma = \frac{\alpha^2(\alpha+\beta)}{\beta(\alpha+2\beta)} \tag{9}$$

Since we know that in equilibrium $\rho^* - E(I|\rho^*) = 0$, we can conclude that

$$\rho^* = \Phi(\sqrt{\gamma}(\rho^* - \bar{r})) \tag{10}$$

It is straightforward to see that the expected payoff is increasing in ρ_i, and therefore it is optimal for player i to adopt the switching strategy around ρ^*. Both players playing these strategies is therefore a Bayesian Nash equilibrium.

4.2.2 Unique equilibrium

In the last section, we concentrated on switching strategies. We will now see that the switching strategy is indeed a unique equilibrium, provided that the noise is small enough. The proof is informal; see Morris and Shin (2001) for a more formal proof.

The first condition for a unique equilibrium is that there is a unique switching strategy, i.e., there is only one solution to (10). This implies that the graph of the scaled cumulative normal should cross the 45°-line only once. Since the slope of the cumulative normal reaches a maximum at the mean, $\bar{r}$, a sufficient condition for the existence of at most one switching strategy equilibrium is that the slope of $\Phi\left(\sqrt{\gamma}(\rho^* - \bar{r})\right)$ is smaller than or equal to one at $\bar{r}$. So,

$$\sqrt{\gamma}\cdot\phi(\sqrt{\gamma}(\rho^* - \bar{r})) \leq 1 \tag{11}$$

Evaluating at $\bar{r}$ gives $\gamma \leq 2\pi$; if γ is small enough, the switching strategy equilibrium is unique. From the definition of γ we can see that when the signal about the state of nature is very precise relative to the uncertainty about the return on investment, that is, when β is large relative to α, γ will get small. So when the noise is small enough, there is a unique switching strategy equilibrium.

It remains to be shown that, provided that there is a unique equilibrium in switching strategies, there are no other equilibrium strategies than switching strategies. This can be done by iterated deletion of dominated strategies. Note that if r is negative, it is never optimal for player i to invest, no matter what the other player does. Therefore, if ρ_i is sufficiently low, refraining from investment is a dominant action. Thus, both investors rule out strategies which invest for signals lower than $\underline{\rho}_1$, the threshold value, and they also rule out strategies of each other which invest for signals lower than $\underline{\rho}_1$. Now, the most "optimistic" strategy that is left is a switching strategy around $\underline{\rho}_1$, i.e., a strategy that always invests for $\rho_i > \underline{\rho}_1$. The best response to this strategy is to play a switching strategy around $\underline{\rho}_2$, where $\underline{\rho}_2$ solves

$$u\left(\underline{\rho}_2; \underline{\rho}_1\right) = 0 \tag{12}$$

where $u(\rho; \rho^*)$ is the expected utility from investing conditional on one's posterior belief ρ when the other investor follows a switching strategy around ρ^*. To see that this is a best response, note that the expected return on investment is increasing in the investor's own posterior belief. But then, investment cannot be optimal for a signal below $\underline{\rho}_2$. Furthermore, no player will believe that the other player will play the switching strategy around $\underline{\rho}_1$, but rather a more cautious one. The incidence of investment by the other player will thus be lower than that implied by the switching strategy around $\underline{\rho}_1$. So every strategy that invests for $\rho_i < \underline{\rho}_2$ is dominated. Since $u(\cdot;\cdot)$ is increasing in its first argument and decreasing in its second argument, i.e., the posterior belief of the other player, $\underline{\rho}_2 > \underline{\rho}_1$. So the second round of deletion narrows the possible equilibrium strategies down to strategies that invest only if $\rho_i > \underline{\rho}_2$. Through iterated dominance, any strategy that invests $\rho_i < \underline{\rho}$ is deleted, where $\underline{\rho}$ is the smallest solution to the equation $u(\rho;\rho) = 0$. A similar argument can be derived for sufficiently high values of ρ_i. Again, through iterated dominance any strategy is removed that refrains for $\rho_i > \bar{\rho}$, where $\bar{\rho}$ is the largest solution to the equation $u(\rho;\rho) = 0$. But since there is only one solution to the equation $u(\rho;\rho) = 0$, the smallest and the largest solution are the same and there remains precisely one strategy after iteratively eliminating all dominated strategies.

Note that this is a very strong result. As we saw in section 4.1, in the full information case, there are multiple equilibria. Here we see that introducing uncertainty about the state of nature and about the beliefs of the other investor leads to a unique equilibrium. This result is independent of the degree of uncertainty; even when the signal about r becomes infinitely precise, there is a unique equilibrium.

We also saw in section 4.1 that the Pareto efficient outcome is that both players play a switching strategy around $\rho_i = 0$. In the unique switching strategy equilibrium that we analyzed, ρ^* is not equal to zero. Indeed, we can think of ρ^* as a measure of inefficiency; the higher ρ^*, the more inefficiency is introduced due to the coordination problem.

5 Bank Runs

The concept of global games can be applied to many economic situations. In this section, we show an application to bank runs. We first analyze a simple model of bank runs, in the spirit of Diamond and Dybvig (1983). Then we show how the concept of global games can be used in this application.

5.1 The Model

There are three dates, $\{1;2;3\}$, and a continuum of consumers. Every consumer is endowed with one unit of the consumption good. Consumption takes place either at date 1 or at date 2. There is measure λ of impatient types who have utility function $u(c_1 + c_2) = \log c_1$ and a measure 1 of patient types who have utility function $u(c_1 + c_2) = \log(c_1 + c_2)$, where c_1 and c_2 denote consumption at date 1 and date 2 respectively. Since the consumers learn of their types only at date 1, the ex ante probability of being an impatient type is $\frac{\lambda}{1+\lambda}$

The consumers can either store the consumption good for consumption at a later date, or they can deposit it in the bank. At date 1, after learning of their types, the consumers who deposited their wealth in the bank must decide wether they will leave their money in the bank or will withdraw the sum that is permitted in the deposit contract.

The bank's portfolio consists of two assets: it may either hold the deposited wealth in cash, which gives a rate of return of 1, or he may invest it in a risky project, with rate of return $R \cdot e^{-l}$ obtainable at date 2. We assume that $0 < \log R < 1$. l is the proportion of the resources invested that is withdrawn at date 1. Early withdrawal is thus costly, in that it reduces the rate of return, i.e., the project is illiquid. When we define $r = \log R$, the rate of return can be written as e^{r-l}. It is assumed that only the bank has access to the investment project.

5.2 Optimal contract under perfect information

A contract is a pair $(c_1; c_2)$, i.e., a promise by the bank to pay c_1 at date 1 if the consumer withdraws and c_2 at date 2 if the consumer does not withdraw. The optimal contract maximizes expected utility

$$\frac{\lambda}{1+\lambda} u(c_1) + \frac{1}{1+\lambda} u(c_2) \tag{13}$$

subject to two constraints:

$$\lambda c_1 + \frac{c_2}{R} \leq 1 + \lambda \tag{14}$$

$$u(c_1) \leq u(c_2) \tag{15}$$

The first constraint is the budget constraint for the bank: on the left-hand side are the expected costs of the contract $(c_1; c_2)$, which must be less than or equal to the total amount of deposits. The second constraint is the incentive compatibility constraint, which requires that patient types will indeed choose to leave their money in the bank. From the budget constraint, it is straightforward to see that $c_1 = 1$ and $c_2 = R$ is a binding solution. For this solution,

$$u(c_1) = 0 < r = u(c_2) \tag{16}$$

so that the incentive constraint is satisfied. In the optimal contract, any depositor can withdraw the whole of their 1 unit deposit at date 1. At worst, a consumer gets his money back on date 1. Therefore, depositing in the bank is a weakly dominant action for every consumer, and, at date 0, all consumers deposit their money in the bank.

Unfortunately, the optimal contract gives rise to multiple equilibria at date 1. At date 1, all the impatient types withdraw their money, so they conform to the optimal contract.

For the patient types, the story is more complicated, however. If a patient consumer expects all other patient consumers not to withdraw, he receives $r > 0$ at date 2 if he conforms to the contract. But if a patient consumer expects all other patient consumers to withdraw ($i.e.; l = 1$), then his utility from not withdrawing is $r - 1 < 0$. So in this case it is optimal for the patient consumer to withdraw. So there are two equilibria: one where everyone conforms to the optimal contract and one where everyone withdraws at date 1. Note the similarity between the outcome of this simple model of bank runs and the outcome of the coordination game in section 4.1. As in section 4.1, there are multiple equilibria, with desirable and undesirable outcomes, and it is impossible to tell which one will prevail. Therefore, the nature of a bank crisis is completely random and there is no correlation between a bank crisis and the underlying state r.

5.3 Imperfect information

Now we change the information structure of the above model and introduce uncertainty about the log return parameter r. Suppose that r is a random normal variable with mean $\bar{r}$, and variance $1/\alpha$. Furthermore, $0 < \bar{r} < 1$. Now the consumers no longer have perfect information about r, but they rather receive a noisy signal $x_i = r + \varepsilon_i$, with ε_i normally distributed with mean 0 and variance $1/\beta$, and independent across consumers. It is typically assumed that $\beta \gg \alpha$, which means that there is relatively little noise in the signal x_i[1].

What does an equilibrium look like in this game? At date 0, the fundamentals of the game are known, and nature chooses r. At date 1, the depositor learns of his type and receives a signal x_i about r. Conditional on this signal, he forms his belief ρ about the true state r, and about the beliefs of the other depositors. Based on this information, he decides whether or not to withdraw. So a strategy for a depositor is a function that maps from the realization of the signal to the action set {Withdraw, Not withdraw}. A profile of strategies is an equilibrium if, conditional on the information available to depositor i and given the strategies followed by other depositors, the action prescribed by i's strategy maximizes his expected utility. Observing that the structure of the game is exactly the same as the coordination game described in section 4.2, it is straightforward to show that the unique equilibrium of this game is that everyone plays a switching strategy around ρ^*, which in this context means that everyone withdraws when $\rho < \rho^*$, where

$$\rho^* = \Phi(\sqrt{\gamma}(\rho^* - \bar{r})) \quad (17)$$

[1] Remember from section 4 that this guarantees a unique switching strategy equilibrium.

Each individual depositor will withdraw whenever his posterior belief, which is based on the realization of his signal, is below the threshold value ρ^*, or whenever

$$\frac{\alpha\bar{r} + \beta x_i}{\alpha + \beta} < \rho^* \tag{18}$$

This means that a depositor will withdraw whenever the signal he receives falls below the critical value

$$x^*(\rho^*; \bar{r}) \equiv \frac{\alpha + \beta}{\alpha}\rho^* - \frac{\alpha}{\beta}\bar{r} \tag{19}$$

Since $x_i = r + \varepsilon_i$, the proportion of depositors that withdraw is a function of r, and is given by

$$l(r) = \Phi\left(\sqrt{\beta}\,(x^*(\rho^*; \bar{r}) - r)\right) \tag{20}$$

This can easily be seen by recognizing that $l(r)$ is the proportion of the depositors whose posterior belief is below ρ^*.

5.4 Implications

According to the Diamond and Dybvig model, the incidence of withdrawal is uncorrelated with fundamental values like the return r, and is instead completely random. However, empirical studies suggest that there are correlations between fundamentals and bank runs. This evidence should not lead to a full rejection of the Diamond-Dybvig story, because it sheds light on the coordination problem that seems to play a crucial role in the onset of bank runs. The above analysis provides a third way of looking at bank crises. As can be seen from (20), the fundamentals do matter. l is decreasing in both r and ρ^*, which means that a higher rate of return does reduce the incidence of withdrawal. But fundamentals are definitely not the whole story. The banking panic is self-fulfilling in the sense that individual investors only withdraw because they expect others to do so. Thus, the theory suggests both that bank runs are correlated with poor fundamentals and that inefficient self-fulfilling panics occur.

6 Creditor Coordination and the Price of Debt

Creditors face a coordination problem when facing a borrower in financial distress. A creditor may be tempted to foreclose on the loan or seize any assets it can, fearing similar actions by other creditors. This fear is self-fulfilling, in that the foreclosure of loans or the liquidation of assets increase the probability that the project fails. Through this coordination problem, viable projects may fail. Therefore, this coordination problem will have an effect on asset pricing. However, most theories of asset pricing do not explicitly deal with coordination problems. One obvious reason for this is that coordination problems lead to multiple equilibria, as we saw above.

There is also empirical evidence supporting the importance of coordination problems. For example, in the financial market turbulence of 1998, which was a period of extremely high yield spreads, access to bank lending was much less affected than access to the bond market. The solution to this empirical puzzle may be that bank lending, with a small number of creditors, suffered much less from the coordination failure between creditors than lending through the bond market. There is also empirical evidence suggesting that the classical way of valuing debt, by using the Merton (1974) model, yields predictions of corporate bond prices which are higher than the actual observed prices, and that the prediction error is higher for riskier bonds.

In order to incorporate the coordination problem into a theory of asset pricing, we need to know the incidence of coordination failure as a function of the fundamental values underlying the asset. Once we know this, we can in corporate the risk of coordination failure into the price of the debt.

6.1 The Model

There is a group of creditors, indexed by the unit interval $[0; 1]$, financing a project with an uncertain payoff V, which is realized at the final stage. Each individual creditor is small, i.e., the individual stake is negligible as a proportion of the whole. The face value of repayment at the final stage equals 1. At an interim stage, the investors have an opportunity to review their investment. They can choose either to roll over the loan until maturity or to foreclose at the interim stage. Foreclosure gives a payoff λ, $0 < \lambda < 1$, which can be thought of as the value of selling collateral. The payoff to rolling over depends on whether the project will succeed or not. This is influenced by two factors: the underlying state θ and the proportion l of creditors who foreclose. The realized value of the project is given by

$$V(\theta; l) = \begin{cases} 1 \text{ if } zl \le \theta \\ 0 \text{ if } zl > \theta \end{cases}, \tag{21}$$

where $z > 0$ is a parameter which measures the severity of the disruption caused by foreclosure. Suppose the creditors have perfect certainty about θ. Note that if $\theta > z$ the project will succeed, and the creditors will always choose to roll over the debt, irrespective of what the other creditors do, i.e., irrespective of l. Conversely, if $\theta < 0$, it is always better to foreclose at the interim stage. However, when $\theta \in (0; z)$, whether rolling over or foreclosure is optimal depends on l, that is, on what the other creditors choose. Therefore, there is a coordination problem when $\theta \in (0, z)$. Note that this coordination problem is akin to the bank run problem described in section 5.2. So, again, we end up with multiple equilibria.

6.2 Imperfect Information

As in the bank runs model, we proceed by assuming imperfect information. At the beginning of the game, the creditors know that θ is a normally distributed random variable with mean y and variance $1/\alpha$. At the interim stage, the creditors receives information concerning θ, in the form of the realization of a noisy signal $x_i = \theta + \varepsilon_i$, where the ε_i's are independent and normally distributed with mean 0 and variance $1/\beta$. Based on this signal, creditor i's posterior belief about θ is normally distributed with mean

$$\xi_i = \frac{\alpha y + \beta x_i}{\alpha + \beta} \tag{22}$$

and variance $1/(\alpha + \beta)$. Now, every creditor adopts a strategy, which tells him whether to roll over or to foreclose, conditional on his posterior. We want to find an equilibrium, that is a strategy for each creditor, such that following this strategy maximizes a creditor's payoff conditional on his signal and given that the other creditors play the strategies in the profile. Analogous to the analysis of section 4, it can be shown that there is a unique equilibrium, provided that the noise is small enough.

First, we define $U(\xi)$, the payoff to rolling over given ξ, if all other creditors follow a switching strategy around ξ. Note that there is an equilibrium in switching strategies around ξ if ξ solves $U(\xi) = \lambda$, that is, the expected conditional payoffs of rolling over and foreclosure are equal. Furthermore, using the iterated dominance argument that we used in earlier sections, we can show that if there is a unique switching strategy equilibrium, then there is no other equilibrium. For a switching strategy to be optimal, the payoff of rolling over must be increasing in ξ for all ξ, $U'(\xi) \geq 0$. Note that this is also a sufficient condition for uniqueness of the switching strategy equilibrium. It can be shown that $U'(\xi) \geq 0$ for all ξ if and only if[2] $\alpha/\sqrt{\beta} \leq \sqrt{2\pi}/z$. So when this condition is met,

[2] Observe that

$$U(\xi) = \int_{\psi}^{\infty} \phi\left(\sqrt{\alpha+\beta}(\theta - \xi)\right) d\theta = 1 - \Phi\left(\sqrt{\alpha+\beta}(\psi - \xi)\right),$$

where $\phi(.)$ is the density of the standard normal and ψ is the failure point.

$$U'(\xi) = -\sqrt{\alpha+\beta} \cdot \phi\left(\sqrt{\alpha+\beta}(\psi - \xi)\right) \cdot \left(\frac{\partial \psi}{\partial \xi} - 1\right)$$

there is a unique equilibrium.

Now we want to solve for this unique equilibrium. If l is determined by everyone following the switching strategy around ξ, what is the critical value of θ for which the project succeeds? We call this critical value of θ the failure point. Note that the failure point depends on the switching point ξ, and that the switching point depends on the failure point. The failure point is the point that solves

$$\theta = zl \tag{23}$$

Conditional on state θ, x is normally distributed with mean θ and variance $1/\beta$. Therefore, if all creditors follow the switching strategy around ξ, the proportion of the creditors who have a signal that is lower than the critical value is

$$l = \Phi\left(\sqrt{\beta}\,(x(\xi; y) - \theta)\right) \tag{24}$$

Substituting this into (23) gives an expression for the critical value of θ at which the project succeeds, the failure point ψ:

$$\psi = z\Phi\left(\sqrt{\beta}\,(x(\xi; y) - \psi)\right) \tag{25}$$

We also know that $U(\xi) = \lambda$, which we can rewrite[3] as

$$1 - \Phi\left(\sqrt{\alpha + \beta}\,(\theta - \xi)\right) = \lambda \tag{26}$$

From the last two equations we can solve for the failure point

$$\psi = z\Phi\left(\frac{\alpha}{\sqrt{\beta}}\left(\psi - y + \Phi^{-1}(\lambda)\frac{\sqrt{\alpha + \beta}}{\alpha}\right)\right) \tag{27}$$

6.3 Implications

Note that $\psi > 0$. Since the efficient outcome is $\psi = 0$, there is an efficiency loss due to the coordination problem. We can also see that the failure point depends on the parameters of the problem. One important conclusion is that failure occurs at higher values of the fundamentals when y, the mean of θ, is low, that is, when the debt is of low quality.

To see the consequences of this model, let's look at the value of an unsecured loan. The owner of such an asset only receives a positive payoff when the true state is higher than the failure point ψ. The ex ante price of such a loan given y is

$$W(y) = \int_{\psi}^{\infty} \phi\left[\sqrt{\alpha}\,(\theta - y)\right] d\theta = 1 - \Phi\left[\sqrt{\alpha}\,(\psi - y)\right] \tag{28}$$

Hence, $U'(\xi) \geq 0$ iff $\frac{\partial \psi}{\partial \xi} \leq 1$. Totally differentiating (25) gives the result.

[3]See footnote 2.

The effect of a shift in the ex ante mean y is thus given by

$$\frac{dW}{dy} = \sqrt{\alpha}\cdot\phi - \frac{\partial\psi}{\partial y}\sqrt{\alpha}\cdot\phi \tag{29}$$

The first term is the "conventional" effect: if y decreases, the distribution shifts to the left, and, with ψ fixed, this increases the left-hand tail. So a decrease in the fundamental value increases the downside risk of an asset. However, the second term is the novel feature. It arises from the fact that a shift in y also causes ψ to shift. For high quality loans, i.e., for high y, lower realizations of the signal will still lead to a successful outcome of the project, that is, it will lower the failure point. Since $\partial\psi/\partial y < 0$, the second effect reinforces the first effect. This second effect is due to the coordination problem.

For the creditor, a deterioration in the fundamental values in terms of a fall in y implies that the asset value of the loan is falling at a rate more than proportional to the fall in y. So by neglecting the coordination effect, the creditor is overestimating the value of his asset or of his portfolio. This theory thus can explain the evidence against the predictions of the Merton model. The Merton model takes into account only the first effect, and will therefore underestimate the downside risk of an asset and predict too high prices. Furthermore, we observed that this prediction error is increasing in the riskiness of the asset. This feature is also consistent with this theory, because it predicts a higher failure point for lower quality assets, i.e., for lower y. This means that for lower quality debt, the coordination problem leads to more inefficiency and thus to a lower price.

References

Carlsson, H. and E. van Damme (1993), Global games and equilibrium selection. Econometrica 61, 989-1018.

Diamond, D. and P. Dybvig (1983), Bank runs, deposit insurance and liquidity. Journal of Political Economy 90, 401-19.

Krugman, P. (1979), A model of balance of payments crises. Journal of Money, Credit and Banking 11, 311-25.

Merton, R.C. (1974), On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance 29, 449-470.

Morris, S. and H.S. Shin (1998), Unique equilibrium in a model of self-fulfilling currency attacks. American Economic Review 88, 587-97.

Morris, S. and H.S. Shin (2001), Coordination risk and the price of debt. mimeo.

Obstfeld, M. (1996), Models of currency crises with self-fulfilling features. European Economic Review 40, 1037-47.

Economics of Transition
Gerard Roland

Report by Mindel van de Laar, Maastricht University[1]

1. Introduction

The transition from socially planned economies towards market economies, as has happened in the countries of Central Europe, Eastern Europe and the Former Soviet Union so distinctively during the last decade, can be considered one of the major economic developments of the last century. The key challenges of transition, according to the Washington consensus, are reforms in the field of Liberalization, Privatization and Stabilization. With liberalization is mainly meant freeing the prices, domestic and foreign trade and investments and elimination of state control (as in liberalizing markets). Privatization mainly involves privatizing of state owned firms and property. Stabilization is focusing on macro-economic factors, as for instance reducing inflation and unemployment, and balancing the national accounts.

More recently two important aspects are often added to these 3 key challenges of transition, being Institutionalization and Social Norms. The first years of transition these two aspects were not considered as main priorities, but currently both the transformation of institutions and the social norms of the society are considered crucial for a successful shift to a market economy.

In this summary I will focus on 2 main aspects of transition economies, being the choice of transition reform strategy and the importance of soft budget constraints during the reform period.[2]

[1] BIRC, PO-Box 616, 6200 MD Maastricht, Email: M.vandelaar@algec.unimaas.nl

[2] *Transition report 1999* (EBRD, 2000), Zecchini (1997).

2. Big bang or gradual reforms

2.1 The 2 main reform strategies: shock and gradual reforms

At the start of transition, countries were facing the difficult choice to decide on what reform strategies to follow. No country had at that time undergone such a swift and all-compassing regime change as the one they were facing from a socialist to a market economy. The only comparable event would probably be their own changes during the shift from market economies to socialist economies in the early 1900's and after the Second World War. As such there was no clear example on how to order the reforms and at what speed. Also no expertise was available at national or international level to guide the transition process.

After a decade of transition can be concluded that 2 different approaches of reform were dominating, being the big bang or shock approach and the gradualist approach. The shock approach (or all out approach) aims at transforming the whole economy in one wave of reforms, including aspects like price and trade liberalization, opening up markets, privatizing state owned firms and imposing stabilization policies to control inflation. The basic idea of the shock approach is to minimize the time of reform, reshape the country quick so it can start growing and become productive again. This way the disadvantages and loss of the reforms by the population are kept moderate, and reforms are undertaken while there is still a significant support for them among the population. A shock approach does not mean all reforms are done quickly. Several reforms are always long-term reforms, but the initial starts of the reforms are all at the same time. Poland is one of the countries that followed a rapid reform process in 1990, as well as East Germany and several NIS states (Newly Independent States).

The gradual approach or piecemeal approach starts with smaller scale reforms that are expanded once they have proven successful. A few necessary but hard reforms will be taken immediately, but afterwards the reform will be a slowly ongoing process. The basic idea of the gradualist way of thinking is to spread the reforms over time, and as such also limit the negative effects of the reforms for the population. The main example of a radualist reforming country is China, and less distinctive examples are Hungary and Slovenia[3].

[3] *Transition report 1999* (EBRD, 2000)

The split up of countries in gradual or big bang approach is not as clear-cut as it might seem. For example Poland that is mentioned often as an example of a country that opted for a big bang approach has also implemented some gradual reforms in specific areas. And the gradual reformers in Europe did liberalize their whole economies at once, which is characteristic for a shock approach. The choice of following either one of the approaches depends on many factors, such as timing of the first reforms (those countries that started reforming before 1989 chose a gradual approach), political support of the reforms and economical situation at the start of transition.

2.2 Reform under uncertainty: a model

The choice options whether or not to engage in reform, can be described using a 2 period model[4]. Ex-ante people can accept or reject reforms. The payoff of the reform is equal to g, with $g>0$, when the reform is positive for an individual. If a reform has negative consequence, the payoff is equal to l, with $l < 0$. In general the population will favor the reforms ex-post if the population benefiting of the reform p is larger than the population with negative outcome $(1\text{-}p)$[5]. However, people are assumed to be risk-neutral ex-ante and with no reform the payoff is assumed to equal 0. So ex-ante people will only vote in favor of reform as long as the discounted expected payoff in total $(1\text{-}d)(pg + (1\text{-}p)l) > 0$, where d is the discount rate. This situation is visualized in figure 1

Figure 1:

Period 0	**reform**	**Period 1**
	p →	payoff g if benefit from reform, with $g > 0$
payoff no reform =0		
	$1\text{-}p$ →	payoff l if lose from reform, with $l < 0$

[4] Fernandez and Rodrik (1991).
[5] According to majority voting the population will favor reform ex-post if $p > 1/2$.

The next extension of the model is the inclusion of a cost of reform reversal. At time t=0 the population will have to decide in favor or against a reform. But in addition, at time t=1 the population is assumed to vote whether or not they would like to reverse the reforms. The costs of reform reversal will equal E, with $E > l$. This is known ex-ante, and will influence the voting behavior of the population negatively. Instead of comparing the expected payoff of reform $pg + (1-p)l$ with the status qua payoff of 0 (no reform), the will now compare it with the costs of reform reversal. The population will only vote in favor of reform if the discounted expected payoff of reform $(1-d)(pg + (1-p)l) > E$.

According to Roland and Dewatripont (1995)[6] this bias against reform might be solved by means of a gradual reform strategy. In a big bang approach, all reforms start at the same time, assuming the final results will be beneficial for the whole economy. The advantages are that there are certain complementarities of the reform[7], so the payoff from all reforms together will be higher than the sum of the payoffs of all separate reforms. However, the costs of reversal are also higher than under gradualist reforms, and there is no possibility to learn from previous reforms. Besides, the losers from the reform will suffer a lot during the period of reform.

When opting for a gradual reform process, the country implicitly chooses for a smaller (or even negative) payoff and also delays several reforms, which is considered negative. However, it does hold the option to reverse the reforms at an early stage if the first reforms are not considered positive, and in that case the costs of this reversal are still relatively low. Besides, the country has the possibility to learn from the gradual reforms, which influences the payoff positively, and the negative consequences of reforms are limited and spread over a longer period of time.

So basically gradualism can be more optimal, as long as the information obtained from the first reforms and the option value of waiting exceed the loss of complementarities and the reversal costs of the partial reforms are low. On the other hand a negative consequence it that several unpopular reforms might not be undertaken under

[6] Roland and Dewatripont (1995).

[7] For example institutions need to build up only once, all votes can be included in one voting round, and information provision can include information of all reforms at once.

gradualism, that would have passed when they would have been included in a big bang reform package[8].

2.3 *Strategies of sequencing reforms: an example*

Considering all these factors that are influencing whether or not a reform or reform package is accepted, it is reasonable to understand there are several political strategic ways to best implement reforms. Assume there are 2 reforms, and there are winners and losers of both reforms. As long as there are no complementarities, and no reversal costs, both shock and gradual reforms will continue as long as the expected payoff exceeds 0. When there are reversal costs and no complementarities, a big bang reform package will only be accepted if they expected payoffs of all reforms exceed the reversal costs. The same goes for a gradual approach, though in that case the reversal costs are significantly lower, which results in a higher probability the gradual reforms are accepted than the shock reforms. When there are no reversal costs, but there are complementarities, a gradual approach is always inferior to a big bang approach, due to the loss of complementarities. When there are both complementarities and reversal costs, several scenarios are possible. These scenarios are explained in the following example:

There are 2 independent reforms: 1 and 2. There is a probability p for in individual to gain in each reform, and of (1-p) that it will lose in the reform. The reversal costs equal E, and the complementarities equal y. The payoffs, ex-ante, will be divided as follows:

	gain reform1	lose reform 1	gain reform 2	lose reform 2
payoff amount	g	-2g	2g	-g
proportion of the population that will receive that payoff	2/3	1/3	1/3	2/3

In a big bang approach, the reforms will be accepted as long as the total payoff exceeds the reversion costs. In this example that is the case as long as $-y > -E$[9]. With gradual reforms, with majority voting, the situation depends on the sequencing of reform. If reform 1 is first implemented, 2/3 of the population will have gained from that

[8] Mathematically the trade off can be modeled. This is extensively described in Roland and Dewatripont (1995).

reform and will vote for continuation as long as the expected payoff from the second reform exceed the cost of reversal, which is as $0 > -E$[10]. 1/3 Of the population lost from reform 1, and might vote for reversal if the loss from the reform plus their expected payoff from reform 2 is still resulting in a bigger loss than having to pay the costs of reversal, so if
$(-2g) > -E$[11]. In total in this scenario therefore at least 2/3 of the population will vote in favor of the reform, since $-E$ is always negative. If you would start with reform 2, only 1/3 of the population will gain from the first reform, and vote in favor of the second one as long as $0 > -E$. The other 66 % of the population will only advocate the second reform as long as $-g > E$. The chance that the second reform (being reform 1 from the table) is not voted in favor for is therewith higher, since the majority of the population will have lost already[12]. Naturally, in the end it does not matter in terms of gaining and losing which reform is implemented first, but in terms of sequencing the chance that both reforms will be voted in favor of is higher when starting with reform 1. This kind of sequencing issues becomes even more important if there are possibilities for compensation between winners and losers.

In practice the split up of Czechoslovakia might be explained by similar voting behavior. Also the fact that the reforms on democracy issues, that were often popular among the population, were undertaken early in the reform process, with afterwards the less popular economic reforms can be considered a well thought of form of sequencing.

---

[9] Calculation: $2/3\ (g) + 1/3\ (-2g) + 1/3\ (2g) + 2/3\ (-g) - y > -E$

[10] 2/3 of the population gained at reform 1 and will vote in favor of reform 2 as long as $1/3\ (2g) + 2/3\ (-g) > -E$

[11] 1/3 of the population lost from reform 1 and will vote in favor of reform 2 as long as $-2g + 1/3\ (2g) + 2/3\ (-g) > -E$

[12] 2/3 of the population lost at reform 2 and will vote in favor of reform 1 as long as $-g + 2/3\ (g) + 1/3\ (-2g) > -E$, so if $-g > E$. Only 1/3 of the population gained at reform 2 and will vote in favor of reform 1 as long as $2/3\ (g) - 2g + 1/3\ (-2g) > -E$, so if $0 > -E$

3. Soft Budget Constraints

The second part of this summary paper will deal with the issue of soft budget constraints, and the relevance of these constraints for transition economies.

3.1 What are Soft Budget Constrains

In a socialist economy, price mechanisms were largely absent. The economy was mainly supply driven, and the government set the prices as a fixed mark-up over costs. The output of each firm was determined by the central plans. This resulted in an enterprise sector that was focused on achieving the agreed output, neglecting the costs and efficiency of production and necessity of the demand for the goods in the market. In addition, the government also supported firms when they were not functioning successfully by means of tax relieves or subsidies. As a result profits were not displaying any information, and the government bailed out all firms that were operating with losses. Naturally, this resulted in large distortions in the enterprise sector and as such the whole economy. The most important were; a too large industrial sector; too few small and medium sized enterprises; the goods supplied were not necessarily those that were demanded by the market; firms had excessive inventories and the goods were of low quality[13].

As a result, the business culture in the countries that started transition was not based on market principles like payment in time and profitability. Within the enterprise sector, the soft budget constraints (SBC's) were maintained. Just like in the socialist times, firms continued to supply each other goods without asking payment in cash. Large inter-enterprise arrears were building up and banks and governments were more or less forced to continue bailing out the firms in order to prevent the whole sector to collapse. Removing the soft budget constraint of firms was considered one of the important reforms needed in order to succeed in transition to a market economy, but how this was best done in practice was not clear at all.

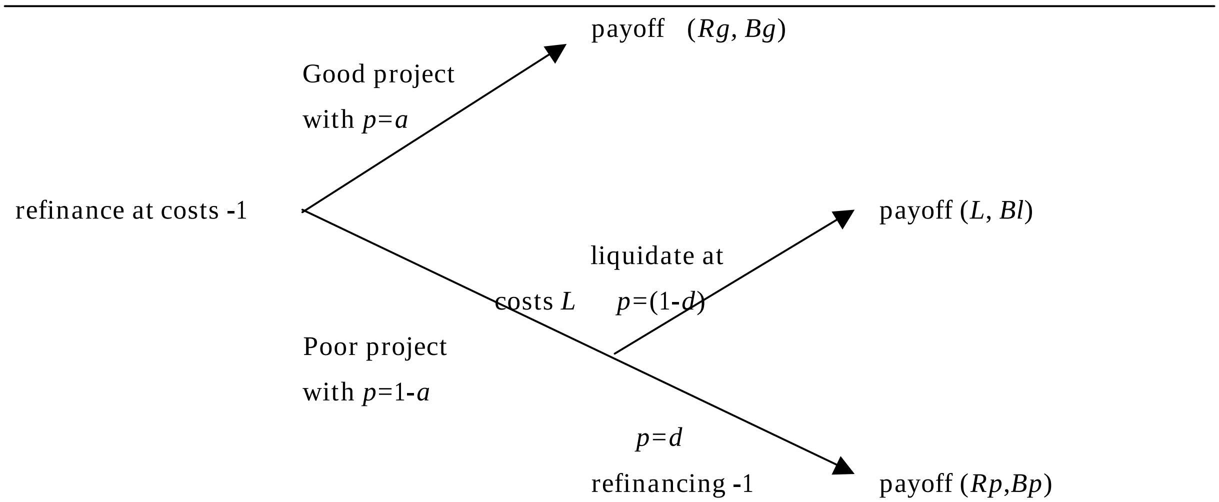
[13] Berg (1994).

3.2 Soft Budget Constraints: a model

Given that under socialism almost all firms were operating under soft budget constraints, it was impossible for the governments to bail out all enterprises at the start of transition. The amounts involved were so large that governments simply did not have enough means to cover all the inter-enterprise arrears and other debts of the state owned enterprises. As a result the soft budget constraints remained present in the early years of transition, and had a devastating effect on the economy. Badly functioning firms were not given any incentive to restructure their enterprise as long as SBC's constraints remained present. Even privatizing the banking sector did not solve the problem, since banks inherited the portfolio of their Former State Owned bank, including all the bad debts of the enterprises. They might have intending to function under hard budget constraints, as regular commercial banks in market economies do. But at the same time they were often facing the difficult choice to refinance bad debt, hoping that some of the earlier debts might be repaid once the firm becomes profitable, or stop financing the firm. If the bank would not refinance the firms, they might go bankrupt and all previously provided credits would never be repaid. This situation can be modeled the following way:

figure 2

period 1 **period 2**



At the start of period 1 the bank can decide whether or not to refinance a firm, that is with probability a a good performing firm, and with probability (1-a) a poor performing firm. In general banks will decide to refinance all firms as long as the total expected payoff of the credit exceed the initial refinancing costs of 1, so if :
$a(Rg+bg) + (1-a)[d(Rp+Bp-1) + (1-d)(L+Bl) > 1$.
After 1 period, the bank will know what kind of firm it has refinanced. There is no problem if the firm turns out to be good functioning. In that case it will achieve good revenues, and will be able to repay the bank at a later stage. However, if the firm turns out to be poorly performing the bank has the option to liquidate the firm in the next period, which will cost L, or to refinance the firm again with a credit of 1. After the 2nd refinancing operation, the firm is expected to get revenue of Rp. In addition to the expected revenues, private benefits are influencing the banks decision. If the firm will be refinanced no layoffs are needed, people receive income and this will translated in private benefits equal to Bp. With liquidation, the private benefits equal Bl, which is most likely equal to zero or very low, since there are no major benefits to be expected when a firm closes.

In general refinancing of the poor performing enterprises will take place as long as the expected payoffs minus the costs of refinancing will be higher than the payoffs of liquidation, so if $L+Bl < Rp+Bp-1$. The initial refinancing costs of -1 are not taken into consideration during this decision since they are seen as sunk costs. However, if ex-ante this problem would have been known, the initial refinancing would not have taken place if $Rp+Bp < 2$. If $L+Bl+1 < Rp+Bp < 2$ the firm will be refinanced ex-post, even though ex-ante this would not have been decided upon.

Poor performing firms thus have an incentive to ask for loans as long as the banks expected payoff ex-ante exceeds the costs of financing the firm. And once financing has been provided to a poor performing firm, the bank will have an incentive to refinance the firm, simply because that is cheaper than liquidating it. As a result banks are keeping soft budget constraints in tact ex post, even though this might be against their intentions ex ante.

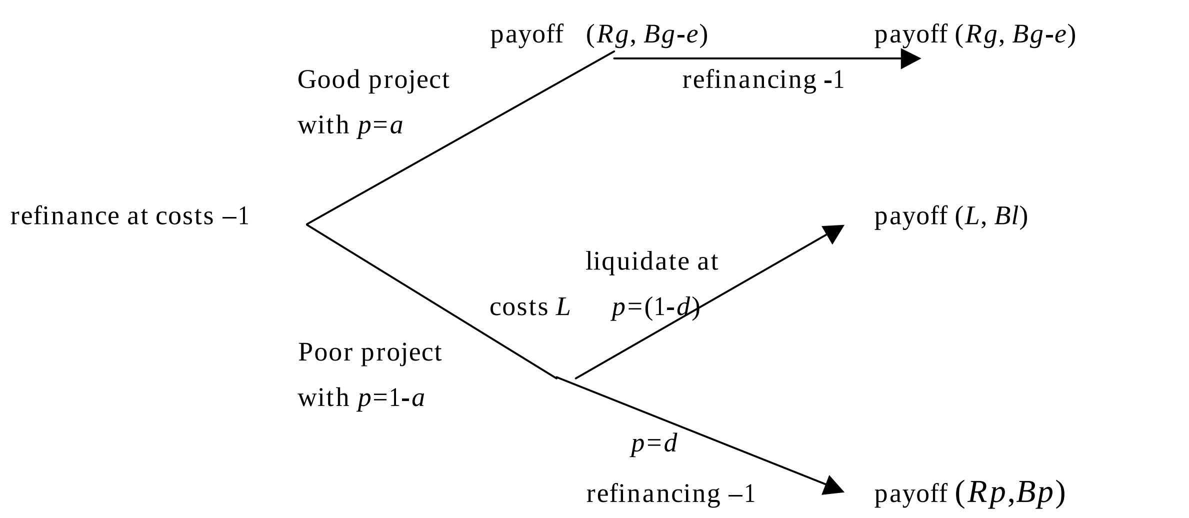
3.3 The Ratchet effect

As a result of refinancing poorly performing firms, good firms in transition countries had a hard time obtaining the necessary financial loans to keep their business operational. This is called the Ratchet effect. Assume good firms need refinancing in order to continue their business after 1 period. They also have to put effort in restructuring their enterprise during period 1, which costs e. If a firm puts effort in restructuring, their private benefits will be lower than without putting effort, due to for instance costs of layoffs, stricter working conditions or reduced salaries. However, it is assumed that the total payoff of the good performing firm that shows effort (Re+Be-e) exceeds the payoff of the good firm with no effort (Rg+Bg), so good firms should be encouraged to put effort in restructuring.

But when the firm requests refinancing at period, it has to compete with the poorly performing firm that did not restructure and also needs the financial means to stay operation (though not profitable). So the bank will only refinance the good firm as long as (Rg+Bg-e-1) > (Rp+Bp-1). Naturally this is discouraging managers of good firms to put effort in restructuring.

figure 3

period 1 **period 2**



Including the option of putting effort also for poorly performing firms can extend the model. The additional assumption is that a poorly performing firm that puts in effort in period 1 becomes a good firm at the end of that period. This does not change the model essentially, though it creates a moral hazard problem. The bank does not know which firms are willing to put in effort in making their enterprise successful, and which ones will need to be either refinanced or liquidated. And again the poorly performing firms that do not make any effort to restructure their firm function will retain financial sources from the banks, keeping the SBC's constraint in tact and creating a credit crunch for the viable firms that do need additional finance.

3.4 Ways to eliminate Soft Budget Constraints

Dealing with SBC's constraints is not easy, there does not seem to be a clear-cut solution to them. However, in market economies the phenomenon of bailing out firms is far less common, so it should be possible to indicate at least some factors that could help eliminating the SBC's.

First of all, privatizing State Owned Banks will be a major step forwards in hardening the budgets. State owned banks are, due to their direct relation with the state, taking the private benefits of the population into account. The government is responsible for the social losses of the country, and might be hesitant do declare a large number of firms or sectors bankrupt at the same time. Commercial banks do not have the responsibility to take care for the well being of the population directly, and will only focus at the revenues of an enterprise. Thus they will much earlier be inclined to liquidate a firm, instead of continuing to refinance.

Besides, commercial banks need their loans to be repaid in the end; otherwise the bank itself will not be successful. Therefore it can be expected that they will monitor their customers better, and make more extensive comparisons and analysis before refinancing firms.

Secondly the entry of new firms will harden the budgets. Former state owned firms have to become more competitive and customer-oriented, in order to keep selling their products. Besides, since newly started firms or foreign participants in the market do not have any bad debts with the bank, and do not have to cover costs of restructuring, they might be eligible for loans that might otherwise have gone to the poorly performing firms. And the more firms in the market are competitively operating, the harder it will be to stick to soft budgets. The business environment will change; cash or advance payments will become rule, thus avoiding the building of inter-enterprise arrears.

In practice several policy measures have been taken to deal with the soft budget problems. Several countries opted for government bailouts, where the government at one point took all the bad loans from the banks, thus cleaning their portfolio. This gave banks the opportunity to start working on a credible portfolio and not having to deal with previous mistakes. Also all inter-enterprise debts were sometimes cancelled off against each other, leaving the firms without huge payment arrears. Another solution was swapping debts for assets, giving firms the opportunity to pay off their debts using assets. A fourth policy done was a re-capitalization of banks. Governments provided banks with additional money, leaving them some room to both deal with their previous credits and work on a current portfolio.

4. Conclusions

In 1989, when the first large group of countries in Central and Eastern Europe started the transition process, no one really knew all the aspects it would involve. A decade later it became clear that the process is a slow one, changes are taking time and the countries need time to recover. Though many countries are well underway in transforming themselves into viable market economies, several other countries are still struggling with basic reforms. It became clear that the choice of reform strategy, being gradual of shock, is not a major determinant in the success of the reform package. More important is the commitment of the country to reform, including the less popular but necessary changes. When transition ends and being a developed market economy begins is not at all clear, though several fundamental reforms like liberalization of markets and prices, privatization of state owned property and stabilization (including hardening the budget constraints) are absolutely basic for a market economy.

References

Main Reference:

Roland, Gerard, "***Transition and Economics*** *Politics, Markets, and Firms*." MIT Press, USA, August 2000

Additional References:

EBRD, “*Transition Report 1999: Ten years of transition*”, London, United Kingdom, 2000

Berg, Andrew, “*Does Macro-economic Reform cause Structural Adjustment: Lessons from Poland*”, Journal of Comparative Economics, 18, pp 376-409, 1994

Roland, Gerard, “*Political Constraints and the transition experience*”, S. Zecchini (ed), Lessons from the economic transition, Dordrecht, Kluwer for OECD, 1997

Dewatripont, M. and Roland G., “*The design of reform packages under uncertainty*”, American Economic Review, vol 85 (5), pp. 1207-1223 (1995)

Fernandez and Rodrik, “*Resistance to Reform: Status Quo bias in the Presence of individual specific uncertainty*”, American Economic review , 81 (5), pp. 1146-1155, 1991

Upcoming....

Month	***Day***	***NAKE activity***
June 2001	11 - 15	Workshop at the Vrije Universiteit Amsterdam
September 2001	7	Start of Utrecht courses, block I 2001/2002: 01.44 - Kooreman *The economics of household behaviour* 01.46 - Magnus *Optimisation methods in econometrics* 01.05 - Folmer *Environmental problems and policy: A theoretical introduction* 01.20 - Van Damme *Topics in applied microeconomics* 01.15 - Brenner *Social economics: Heterodox approaches to economic theory* Course descriptions see: **http://few.kub.nl/nake/courses.htm**
October 2001	12	NAKE day 2001
October 2001	26	Start of Utrecht courses, block II 2001/2002
December 2001	10 - 14	Workshop, location to be announced

Information about these activities can be obtained from Marty Roovers, nake@kub.nl