Essays on Weak Identification

by

Zhaoguo Zhan

B.Eng., Renmin University of China, 2003

MSc, London School of Economics, 2005

M.A., Brown University, 2007

Submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in Economics at Brown University

PROVIDENCE, RHODE ISLAND

May 2011

This dissertation by Zhaoguo Zhan is accepted in its present form
by the Department of Economics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date 04/26/2011 _____

Frank Kleibergen, Advisor

Recommended to the Graduate Council

Date 04/26/2011 _____

Blaise Melly, Reader

Date 04/26/2011 _____

Geert Dhaene, Reader

Approved by the Graduate Council

Date _____ _____

Peter M. Weber, Dean of the Graduate School

Vitae

The author was born on November 5th, 1980 in China. He received his Bachelor in Engineering from Renmin University of China in 2003, and MSc in Econometrics and Mathematical Economics from London School of Economics in 2005. He entered Brown University in 2006 to pursue his degree in Economics. He received his M.A. in 2007 and Ph.D. in 2011.

Acknowledgements

I am deeply indebted to my advisor Frank Kleibergen for continuous support and encouragement. I thank Blaise Melly and Sophocles Mavroeidis for inspiration and insightful comments. Thanks to Geert Dhaene for reading the manuscript. I also owe much to my colleagues, Toru Kitagawa, Yuya Sasaki, Alexei Abrahams and Philipp Ketz, who spent time discussing econometrics with me throughout my graduate study. Lastly, I am most indebted to my wife Flora for her consideration and support.

Abstract of " Essays on Weak Identification " by Zhaoguo Zhan, Ph.D., Brown University, May 2011

Recent developments in econometrics have revealed weak identification in empirical studies. The linear Instrumental Variable (IV) regression with weak instruments is an example that has received sizable attention; when instruments are weak, conventional asymptotics fail to function, and empirical results based on conventional asymptotics are unreliable. The problem of weak identification, however, is not limited to the linear IV regression or the Generalized Method of Moments (GMM) framework. In a broad range of economic models, the quality of inference depends on the treatment of weak identification.

In my dissertation, I consider three issues related to weak identification. The first chapter proposes a method to detect whether weak identification exists. The objective of this chapter is to develop an intuitive tool which can be universally used in IV and GMM applications. The second chapter investigates the empirical studies of the Capital Asset Pricing Model (CAPM) and Consumption CAPM, in which various risk factors are suggested to explain the variation in asset returns. I find that irrelevant risk factors may still appear useful in explaining the variation, because of the weak identification problem induced by irrelevant factors. The third chapter targets the ongoing debate on whether technology shocks increase the hours worked. The debate derives from the highly persistent time series of hours, which makes the impact of technology shocks on hours weakly identified. I construct confidence intervals for this impact by adopting an approach robust to weak identification.

CONTENTS

CHAPTER **One**

Detecting Weak Identification by Bootstrap

1.1 Introduction

The Instrumental Variable (IV) regression and Generalized Method of Moments (GMM) are becoming the standard toolkit for empirical economists, and it is now well known that both IV and GMM applications may suffer from the problem of weak identification. An example that has received sizable attention is the linear IV regression with weak instruments studied in Staiger and Stock (1997). When the strength of identification is weak, the finite sample distribution of IV/GMM estimators is poorly approximated by the normal distribution, which further induces the malfunction of conventional inference methods that rely on the property of asymptotic normality.[1] Two approaches co-exist to handle weak identification: the first approach is to use the robust methods in Anderson and Rubin (1949), Stock and Wright (2000), Kleibergen (2002)(2005), and Moreira (2003), which can produce confidence intervals/sets with the correct coverage, regardless of the strength of identification; the second approach that is popular in practice is to rule out weak identification by pretesting. Although identification-robust methods are available, excluding weak identification in IV/GMM applications has practical importance: if identification is not weak, then the rich set of conventional methods is applicable, making statistical inference and economic decisions much easier. For instance, other than its confidence interval, the point estimator of a parameter is usually preferred by policy makers, but it is not consistent or meaningful unless weak identification is excluded.

In this chapter, I propose a method based on bootstrap resampling to detect whether weak identification exists in IV/GMM applications. This method has the unique feature of providing a graphic view of the strength of identification. In the econometric literature, there exists a group of tests on identification that this chap-

[1] See the article by Stock *et al.* (2002) for a survey.

ter is in line with: for example, in the linear IV model, Hahn and Hausman (2002) and Stock and Yogo (2005) provide tests for the null of strong instruments and weak instruments respectively, and the first stage F test with the $F > 10$ rule of thumb proposed in Stock and Yogo (2005) is widely used; in Hansen (1982)'s GMM framework, which nests the linear IV model, the suggested tests include Wright (2002)(2003), Inoue and Rossi (2008), Bravo *et al.* (2009), etc.

The proposed method is illustrated by Figure 1.1. The exact finite sample distribution of IV/GMM estimators is generally unknown, but can be approximated either by the limiting normal distribution or by the bootstrap distribution. When the IV/GMM models are strongly identified, these two approximation methods are both valid, i.e. both the normal distribution and the bootstrap distribution are close to the exact distribution, and in practice, they can be used exchangeably. Consequently, strong identification implies that the bootstrap distribution is not far away from the normal distribution. When the bootstrap distribution is substantially different from the normal distribution, inference based on these two distributions might contradict; in this situation, it is inappropriate to consider the identification strength as strong. As a result, whether or not weak identification exists can be inferred by comparing the bootstrap distribution with the normal distribution.

Since its introduction by Efron (1979), the bootstrap has become a practical tool for statistical inference. The properties of the bootstrap are explained using the theory of the Edgeworth expansion in Hall (1992), and its econometric applications are illustrated in Horowitz (2001). As an alternative to the limiting distribution, the bootstrap approximates the distribution of a targeted statistic by resampling the data, and there is considerable evidence that it performs better than the first-order limiting distribution in finite samples.[2] However, the bootstrap does not always

[2]See, for example, Horowitz (1994).

work well. When the IV/GMM models are not well identified, for instance, the bootstrap is known to be problematic when it is used to approximate the commonly used t/Wald statistic, as explained in Hall and Horowitz (1996). Nevertheless, the fact that the bootstrap fails still conveys useful information: as illustrated by Figure 1.1, the substantial difference between the bootstrap distribution and the normal distribution indicates that it is problematic to approximate the exact finite sample distribution of IV/GMM estimators by the normal distribution; in other words, the identification strength is weak.

Given the above introduction, it is tempting to apply normality tests, e.g. the Kolmogorov-Smirnov test, to examine whether the bootstrap distribution is normal in order to investigate the strength of identification. However, this route is not productive, when normality tends to be rejected as the number of bootstrap replications becomes large. This is due to the fact that, in general, the bootstrap distribution coincides with the normal distribution only at the limit where there are an infinite number of data points. In an IV/GMM application with a finite sample, the bootstrap distribution is not equivalent to the normal distribution, hence the null hypothesis for normality tests that the bootstrap distribution is identical to the normal distribution does not hold in empirical applications. Although normality tests do not work well, empirical researchers can still eyeball the graph of the bootstrap distribution to evaluate the strength of identification, and the test proposed in this chapter does not impose the equivalence of the bootstrap distribution and the normal distribution; only a substantial difference of these two distributions would induce the rejection of strong identification.

The rest of the chapter is organized as follows: weak identification and the bootstrap strategy are illustrated in Section 2; in Section 3, the bootstrap-based method for determining whether weak identification exists is proposed; the linear IV regres-

sion model is used as an example in Section 4, with an application to Card (1995); in Section 5, Monte Carlo results are presented; Section 6 concludes. Although the linear IV model is employed in this chapter for expository purposes, the proposed method can be extended to the more general GMM framework.

Throughout the chapter, the following notations are used: for an m by n matrix A, A_i (sometimes a_i is used when A is a vector) is its i^{th} row, and $P_A = A(A'A)^{-1}A'$, $M_A = I_m - P_A$, I_m is the m by m identity matrix; $vec(A)$ is the column vector containing the column by column vectorization of elements in A; for an object O, O^* is its bootstrap counterpart; $\Rightarrow$ stands for weak convergence; $\xrightarrow{p}$ stands for convergence in probability; $N(\mu, \sigma^2)$ is the normal distribution with mean μ, variance σ^2; $\otimes$ is the Kronecker product; $||.||$ is the Euclidean norm.

1.2 Motivation and Strategy

1.2.1 Weak Identification, Rank Tests

Let θ denote the population parameter of interest in an IV/GMM application, and $\hat{\theta}_n$ is the estimator for θ, e.g. the two stage least squares estimator. The subscript n for $\hat{\theta}_n$ refers to that $\hat{\theta}_n$ is computed by finite sample of n observations. Assuming that the rank condition and other regularity conditions are satisfied (cf., Wooldridge (2002)), the conventional first-order asymptotic theory yields that $\hat{\theta}_n$ is $\sqrt{n}-$consistent, and asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \sigma^2), \text{ and there exists } \hat{\sigma} \xrightarrow{p} \sigma \tag{1.1}$$

As surveyed by Stock *et al.* (2002), the exact finite sample distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ can be substantially different from the normal distribution $N(0, \sigma^2)$, especially when the rank condition is weakly satisfied. The scenario that (1.1) does not provide a good approximation in finite sample applications is known as *weak identification*.

It is helpful to distinguish *weak identification* from *identification/noidentification*: *identification/noidentification* refers to whether θ can be identified at the population level, hence it is unrelated to the sample size n; in contrast, *weak identification* targets that the first-order asymptotic theory may provide a poor approximation in finite samples. When θ is *identified* in the population, it is possible that the first-order approximation of (1.1) under a given sample[3] does not function well, hence *weak identification* may exist even though there is *identification*; the case that θ is *unidentified* is also nested by *weak identification*, since the first-order approximation of (1.1) also breaks down under *noidentification*. Loosely speaking, if the first-order approximation based on (1.1) is poor, then *weak identification* is concerned.

Since the conventional result of (1.1) breaks down under weak identification, it is important to investigate the identification strength in IV/GMM applications. A natural way is to examine the rank condition, as weak identification is more likely to exist when the rank condition is only weakly satisfied. There are rank tests available to serve this purpose. For example, in the linear IV model with a single endogenous regressor, the rank condition corresponds to that the correlation of the endogenous regressor and the instruments is non-zero, and Stock and Yogo (2005) propose the F test to examine this correlation: if the F statistic is greater than the tabled critical values, typically around 10 for small number of instruments, then the rank condition is considered strongly satisfied, and weak identification is excluded.

[3]Even when the sample size is very large, weak identification could exist, e.g. in some regressions of Angrist and Krueger (1991), there are 329000 observations, but their instruments are weak.

The approach of examining the rank condition, however, has some limitations when it is extended to the more general GMM framework: first of all, the rank statistic in the non-linear GMM may depend on the weakly identified parameters, hence could not be consistently estimated, see, e.g. Wright (2003); secondly, it is not clear how large the rank statistic needs to be in order to decide that identification in GMM is not weak, to the best of my knowledge. In the linear IV model, which can be seen as a special case of GMM, although the tabled critical values of the F test are available, they are derived under the i.i.d and homoscedasticity assumptions, hence it is not appropriate to apply them if heteroscedasticity or non-i.i.d. takes place.

Given the importance of detecting weak identification in IV/GMM applications, and the limitations of the rank test, a question naturally arises: is there a tool that does not have these limitations, and is universally applicable to both IV and GMM applications? This chapter suggests that the bootstrap could be the tool.

1.2.2 Test Strategy by Bootstrap

Let $\hat{\theta}_n^*$ denote the bootstrap estimator of θ, and $\hat{\theta}_n^*$ is the counterpart of $\hat{\theta}_n$. $\hat{\theta}_n$ is computed by the data sample of the IV/GMM application, while $\hat{\theta}_n^*$ is computed by resampling the data sample or the model estimated from the data. The bootstrap counterpart of $\sqrt{n}(\hat{\theta}_n - \theta)$ is $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$.

When identification is not weak, together with mild regularity conditions, the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ asymptotically coincides with the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ (cf., Horowitz (2001)):

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \Rightarrow N(0, \sigma^2) \tag{1.2}$$

There are now three objects: the exact distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, the normal distribution $N(0, \hat{\sigma}^2)$, and the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$. By (1.1) and (1.2), both of which hold under mild conditions, these three distributions are asymptotically equivalent.

The proposed bootstrap strategy for detecting weak identification is to compare the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ with the normal distribution $N(0, \hat{\sigma}^2)$, or equivalently, compare the standardized bootstrap distribution of $\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ with the standard normal distribution. Since the difference between these two distributions is negligible when identification is strong, a substantial difference is the evidence against strong identification.

I find that this bootstrap strategy is appealing: (i) it provides a simple and intuitive way to detect weak identification, i.e. empirical researchers can draw the graph of the bootstrap distribution and compare it with the normal distribution to evaluate the strength of identification; (ii) it is universally applicable to IV/GMM applications, i.e. although the data of IV/GMM applications may not be i.i.d. and homoscedastic, there exist various bootstrap methods to construct the bootstrap distribution, e.g. the block bootstrap in Hall and Horowitz (1996) for time series data in GMM, the pair and wild bootstrap used in Davidson and MacKinnon (2010) for heteroscedasticity in IV. By the bootstrap strategy, as long as it is feasible to construct the bootstrap distribution, detecting weak identification reduces to the comparison of two distributions: the bootstrap distribution, and the normal distribution.

Definition 1. *Conditional on the sample of the observed data, the standardized bootstrap estimator* $X = \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ *follows a distribution with c.d.f.* $F(x)$*. From now on,* ***the bootstrap distribution*** *refers to* $F(x)$ *in this chapter.*

Normality tests appear to be the intuitive choice, since comparing the bootstrap

distribution with the normal distribution is proposed to infer the strength of identification. However, applying normality tests here will almost always induce the rejection of normality, and lead to the conclusion of weak identification.

Take the classic Kolmogorov-Smirnov (KS) test for example. With B bootstrap replications, let $X_i = \frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}, i = 1, ..., B$ denote the i.i.d. bootstrapped estimators after standardization. To test the hypothesis that $X_i, i = 1, ..., B$ are B points drawn from the standard normal distribution $\Phi(x)$, the KS statistic is the supremum distance between $\Phi(x)$ and the empirical c.d.f. $\hat{F}(x)$, scaled by the square root of the number of points. As $B \to \infty$, the KS statistic goes to infinity, instead of converging to the Kolmogorov distribution:

$$KS = \sqrt{B} \sup_x |\hat{F}(x) - \Phi(x)| \to \infty$$

where $\hat{F}(x) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}(X_i \leq x) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}(\frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x)$.

This is because the hypothesis $F(x) = \Phi(x)$ does not hold when $n < \infty$. In other words, in empirical applications where the sample size is finite, $F(x)$ differs from $\Phi(x)$, although the difference may not be substantial, e.g. by the Edgeworth expansion in Horowitz (2001), $F(x) - \Phi(x) = O(n^{-1/2})$. Consequently, even when the difference between $F(x)$ and $\Phi(x)$ is minor, as $B \to \infty$, $KS \to \infty$, i.e. the KS test tends to reject normality when the bootstrap replication gets large.

Said differently, the bootstrap distribution is *not identical to* the normal distribution, although it can be asymptotically equivalent to the normal distribution. The KS test is for testing the hypothesis that the bootstrap distribution is *identical to* the normal distribution. For the purpose of examining whether the bootstrap distribution is close to the normal distribution, it is inappropriate to apply this test.

Instead of verifying the equivalence of the bootstrap distribution and the normal distribution by normality tests, this chapter provides a quantitative measure of the difference/distance between these two distributions: if the measure shows that the difference is substantial, then identification is considered as *weak*; on the contrary, if the measure shows that the difference is negligible, then identification is considered as *strong*.

The measure results from the comparison of two confidence intervals (C.I.): the conventional C.I. derived by inverting the t/Wald test, and the bootstrap percentile C.I.. Consider the practical task of constructing a confidence interval for θ, the parameter of interest. The $100(1-\alpha)\%$ C.I. of θ derived by inverting the t/Wald test is written as:

$$C_t \equiv (\hat{\theta}_n - z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta}_n + z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}) \tag{1.3}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\Phi(x)$. For example, when $\alpha = 5\%$, $z_{1-\alpha/2} \approx 1.96$, and the 95% C.I. of θ is approximately $(\hat{\theta}_n - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta}_n + 1.96\frac{\hat{\sigma}}{\sqrt{n}})$.

Alternatively, a C.I. can also be constructed by the bootstrap percentile method: order the bootstrapped estimators $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$, write the ordered sequence as $\{\hat{\theta}_n^{*(i)}, i = 1, ..., B\}$, where $\hat{\theta}_n^{*(i)}$ is the i^{th} smallest of $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$; define $\hat{\theta}^*_{n,\alpha/2}$ and $\hat{\theta}^*_{n,1-\alpha/2}$: $\hat{\theta}^*_{n,\alpha/2} \equiv \hat{\theta}_n^{*(\lceil B\alpha/2 \rceil)}$, $\hat{\theta}^*_{n,1-\alpha/2} \equiv \hat{\theta}_n^{*(\lceil B(1-\alpha/2) \rceil)}$, where $\lceil x \rceil$ denotes the integer ceiling of x. The bootstrapped $100(1-\alpha)\%$ C.I. of θ is:

$$C_b \equiv (\hat{\theta}^*_{n,\alpha/2}, \hat{\theta}^*_{n,1-\alpha/2}) \tag{1.4}$$

The two intervals in (1.3) and (1.4) are asymptotically equivalent when identification is not weak. To see this, equalizing the boundaries of these two intervals

yields:

$$\frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} = -z_{1-\alpha/2}, \ \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} = z_{1-\alpha/2}$$

These two equalities approximately hold if the distribution of $\hat{\theta}^*_n$ after standardization, i.e. subtracting the estimate and dividing by the standard error, is close to the standard normal distribution.

In practice, the above two methods of constructing the $100(1-\alpha)\%$ C.I. for θ are both commonly used. From a practical point of view, no matter which method empirical researchers use, the correspondent intervals need not be substantially different from each other. If these two intervals do substantially differ, then it is difficult to make reliable economic inference: for instance, one interval may include zero while the other one does not, hence decisions of whether θ is significantly different from zero based on the two different C.I.'s could contradict.[4] The difference between the two intervals in (1.3) and (1.4) boils down to the difference between the bootstrap distribution and the normal distribution. If these two intervals are substantially different, it indicates the bootstrap distribution is substantially different from the normal distribution, hence the identification strength is weak.

The idea of comparing alternative C.I.'s to investigate the identification status comes from Wright (2002), where he compares the interval derived by inverting the robust tests with the conventional interval derived by inverting the t/Wald test. Different from Wright (2002), I use the bootstrap to construct a C.I. for comparison; in addition, Wright (2002) provides an identification test, while in this chapter, I target weak identification instead of identification.

Based on the comparison of the two intervals in (1.3) and (1.4), a measure of the

[4]This problem is encountered in the empirical example of Card (1995).

difference between the bootstrap distribution and the normal distribution, as well as a quantitative definition of weak identification, is provided below.

Definition 2. *Define D as the measure of the difference between the bootstrap distribution $F(x)$ and the standard normal distribution $\Phi(x)$:*

$$D \equiv \frac{q_{1-\alpha/2} - q_{\alpha/2}}{2z_{1-\alpha/2}} - 1$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\Phi(x)$, $q_{\alpha/2}, q_{1-\alpha/2}$ are assumed to be the two unique quantiles of the continuous c.d.f. $F(x)$, i.e. $F^{-1}(\alpha/2) = q_{\alpha/2}$, $F^{-1}(1-\alpha/2) = q_{1-\alpha/2}$.

D is the relative difference of the lengths between the $\alpha/2, 1-\alpha/2$ quantiles of the two distributions. $D = 0$ if $F(x) = \Phi(x)$. The deviation of D from 0 indicates the deviation of $F(x)$ from $\Phi(x)$, and hence suggests existence of weak identification.

Definition 3. *Suppose there is a cutoff $\gamma > 0$. The identification strength is considered as* ***weak*** *in this chapter if $|D| > \gamma$, otherwise the identification strength is* ***strong****.*

It is important to set a non-zero γ: if $\gamma = 0$, then identification tends to be considered as *weak*, because $F(x)$ is not exactly normal except in the limit of $n \to \infty$. For the same reason, it is inappropriate to apply normality tests, which impose $\gamma = 0$ under their null hypothesis.

D can be estimated by the relative difference of (1.3) and (1.4):

$$\hat{D} = \frac{\hat{q}_{1-\alpha/2} - \hat{q}_{\alpha/2}}{2z_{1-\alpha/2}} - 1 = \frac{\frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}}{2z_{1-\alpha/2}} - 1 = \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}} - 1$$

Empirical researchers may have a certain tolerance level for D, which is the threshold[5] γ. For example, a researcher may consider it acceptable if $|D| \leq \gamma = 0.25$, i.e. the relative difference of the two C.I.'s is less than a quarter. If $|D|$ goes above this tolerance level, then it is not unreasonable to determine that the identification strength is *weak*.

To summarize, this chapter proposes to use the difference D between the bootstrap distribution $F(x)$ and the standard normal distribution $\Phi(x)$ to evaluate the strength of identification. If $|D|$ is greater than a given threshold, identification is *weak*.

1.2.3 An Alternative

To my knowledge, Stock and Yogo (2005) is the first to provide a quantitative definition of weak instruments/identification. Like Stock and Yogo (2005), the quantitative Definition 3 of weak identification is also practically motivated: from the practical perspective, it is not appropriate to consider identification as strong if the bootstrap and the limiting normal distribution provide substantially different confidence intervals. Unlike Stock and Yogo (2005), Definition 3 is not directly related to the bias of the conventional IV/GMM estimator or the size of the t/Wald test, while Stock and Yogo (2005) define identification as weak if the relative bias of the conventional estimator or the size distortion of the t/Wald test exceeds a threshold. Although quantitatively different, the definition in this chapter is qualitatively similar to the one in Stock and Yogo (2005): both definitions of weak identification indicate that the distributions of the conventional IV/GMM estimators are poorly approximated by the normal distribution, or said differently, the conventional ap-

[5] A choice of the threshold γ is suggested in the later part of Monte Carlo studies.

proximation of (1.1) breaks down.

In this chapter, I follow the idea in Wright (2002) to use D, the relative difference of the lengths of two alternative C.I.'s, to measure the strength of identification. Alternatively, we could follow Stock and Yogo (2005) and define D as the difference of coverage:

$$D^a \equiv F(z_{1-\alpha/2}) - F(z_{\alpha/2}) - (1-\alpha)$$

Note that D^a is zero if the bootstrap distribution $F(x)$ coincides with the standard normal distribution $\Phi(x)$. Following Stock and Yogo (2005), we can use $\gamma^a = 5\%$, and define identification as weak if $|D^a| > \gamma^a = 5\%$.

These definitions of $(D,\ \gamma)$ and the alternative $(D^a,\ \gamma^a)$ are two sides of the same coin: both the difference in length and the difference in coverage indicate the bootstrap distribution and the normal distribution are different, and there is no clear advantage if either definition is used. For the rest of this chapter, I use $(D,\ \gamma)$[6].

Note that *weak identification* refers to the severe disparity between the exact distribution and the normal distribution, while its quantitative definition stated above rests on the disparity between the bootstrap distribution and the normal distribution. In essence, the bootstrap strategy for detecting weak identification is to use the difference between the bootstrap distribution and the normal distribution as a proxy for the the difference between the exact distribution and the normal distribution.

The so-called bootstrap principle (or the bootstrap analogy) in Hall (1992) can help clarify the proposed strategy. The bootstrap principle states that the mapping from the population to the sample (1^{st} mapping) is similar to the mapping from the sample, which is also the bootstrap population, to the bootstrap resample (2^{nd}

[6]The empirical and MC results are found to be similar, when $(D^a,\ \gamma^a)$ is used.

mapping). By this principle, the identification strength in the 2^{nd} mapping is expected to be similar to the identification strength in the 1^{st} mapping. Consequently, the bootstrap strategy for detecting weak identification is to use the identification strength in the 2^{nd} mapping as the proxy for the identification strength in the 1^{st} mapping: the substantial disparity between the bootstrap distribution and the normal distribution corresponds to the weak identification strength in the 2^{nd} mapping, which further suggests the weak identification strength in the 1^{st} mapping, i.e. the disparity between the exact distribution and the normal distribution is also severe.

The advantage of the bootstrap strategy is clear: the bootstrap population is the given sample, hence the identification strength in the 2^{nd} mapping is known or recoverable. The only randomness in this mapping comes from the randomness of bootstrap resampling, and if the number of bootstrap replications is sufficiently large, this randomness is negligible. Once the identification strength in the 2^{nd} mapping is recovered, it is used to infer the identification strength in the 1^{st} mapping, since they are expected to be similar by the bootstrap principle.

1.3 Estimation and Test

With the quantitative definition as well as the advantage of the bootstrap test strategy, detecting weak identification becomes straightforward. By definition, $|D| > \gamma$ implies weak identification. D, the distance between the bootstrap distribution and the standard normal distribution, needs to be estimated.

Draw B i.i.d. observations from $F(x)$ by bootstrap: $X_i = \frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, $i = 1, ..., B$. The bootstrap distribution $F(x)$ can be estimated by the empirical c.d.f. $\hat{F}(x)$ almost

surely:

$$\hat{F}(x) = \frac{1}{B}\sum_{i=1}^{B} \mathbf{1}(X_i \leq x) = \frac{1}{B}\sum_{i=1}^{B} \mathbf{1}(\frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) \xrightarrow{a.s.} F(x)$$

Consequently, D can be estimated almost surely, by the continuous mapping theorem:

$$\hat{D} = \frac{\hat{q}_{1-\alpha/2} - \hat{q}_{\alpha/2}}{2z_{1-\alpha/2}} - 1 \xrightarrow{a.s.} D$$

It would be ideal if B could be made infinity. For the given B realizations from $F(x)$, though B can be arbitrarily large, whether $|D|$ exceeds γ needs to be tested, and a test serving this purpose is presented next.

Assume the following conditions hold for an IV/GMM model with the conventional estimator $\hat{\theta}_n$, associated with standard error $\frac{\hat{\sigma}}{\sqrt{n}}$:

Assumption 1. *There exist $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$, the i.i.d. draws of the bootstrapped estimator $\hat{\theta}_n^*$; conditional on the sample, the standardized random variable $X = \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ has a continuous density function $f(x)$ that is non-zero in a neighborhood of the two quantiles $q_{\alpha/2}, q_{1-\alpha/2}$, and can be consistently estimated by the non-parametric kernel estimation: $\hat{f}(x) \xrightarrow{p} f(x)$.*

Comments:

1. Under Assumption 1, the joint distribution of the two quantile estimators, namely, $\hat{q}_{\alpha/2} = \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, $\hat{q}_{1-\alpha/2} = \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, is asymptotically normal conditional on the sample, as $B \to \infty$ (see David and Nagaraja (2003)):

$$\sqrt{B}\begin{pmatrix} \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - q_{\alpha/2} \\ \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - q_{1-\alpha/2} \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right)$$

where

$$\Omega = \begin{pmatrix} \frac{(1-\alpha/2)\alpha/2}{f(q_{\alpha/2})^2} & \frac{(\alpha/2)^2}{f(q_{\alpha/2})f(q_{1-\alpha/2})} \\ \frac{(\alpha/2)^2}{f(q_{\alpha/2})f(q_{1-\alpha/2})} & \frac{(1-\alpha/2)\alpha/2}{f(q_{1-\alpha/2})^2} \end{pmatrix}$$

2. Silverman (1998) provides high level assumptions for the consistency of the non-parametric kernel density estimator, while $\hat{f}(x) \xrightarrow{p} f(x)$ is directly assumed here for simplicity. As $\frac{\hat{\theta}^*_{n,\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \xrightarrow{p} q_{\alpha/2}$, $f(q_{\alpha/2})$ is consistently estimated by $\hat{f}(\frac{\hat{\theta}^*_{n,\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}})$; similarly, $f(q_{1-\alpha/2})$ is consistently estimable. The covariance matrix Ω is thus consistently estimable: there exists $\hat{\Omega} \xrightarrow{p} \Omega$. The normal kernel and Silverman's rule of thumb for choosing the bandwidth are used in the empirical application and simulation studies of this chapter.

Theorem. *Under Assumption 1, and conditional on the sample, the following result holds as* $B \to \infty$*:*

$$\sqrt{B}(\hat{D} - D) \quad \Rightarrow \quad N(0, \frac{\Omega_{11} + \Omega_{22} - 2\Omega_{12}}{4z^2_{1-\alpha/2}}) \tag{1.5}$$

where $\Omega_{i,j}$ *is the element of* Ω *at row* i*, column* j*.*

The quantitative definition of *strong* and *weak* identification implies the following decision rule: reject the null of *strong* identification when $|D| > \gamma$. There are two cases that would induce the rejection, $D > \gamma$ and $D < -\gamma$. Consequently, *strong* identification is rejected when $\hat{D}$ is significantly greater than γ (Case 1), or significantly less than $-\gamma$ (Case 2) in the test statistics below.

Case 1: *strong* identification is rejected at 5% if

$$b_1 = \sqrt{B}\frac{\hat{D} - \gamma}{[(\hat{\Omega}_{11} + \hat{\Omega}_{22} - 2\hat{\Omega}_{12})/(4z^2_{1-\alpha/2})]^{1/2}} > z_{95\%}$$

Case 2: *strong* identification is rejected at 5% if

$$b_2 = \sqrt{B}\frac{\hat{D}+\gamma}{[(\hat{\Omega}_{11}+\hat{\Omega}_{22}-2\hat{\Omega}_{12})/(4z^2_{1-\alpha/2})]^{1/2}} < -z_{95\%}$$

Combining these two cases, reject *strong* identification at 5% if

$$|\hat{D}| > \gamma + z_{95\%}\sqrt{\frac{\hat{\Omega}_{11}+\hat{\Omega}_{22}-2\hat{\Omega}_{12}}{4Bz^2_{1-\alpha/2}}}$$

As $B \to \infty$, the test above ends up with a rule of thumb: reject *strong* identification if $|\hat{D}| > \gamma$, hence this test can be substituted by the rule of thumb under sufficiently large B. From now on, this test is referred to as the b test, since it is based on the bootstrap.

1.4 IV and Bootstrap

In this section, a linear IV regression model is used as a specific example to further illustrate the bootstrap approach for detecting weak identification, with an application to Card (1995). Most of the analytical results are well known, for example, the convergence results in (1.1) and (1.2) under mild conditions, and proofs of the listed results are attached in the appendix. The main objective of this section is to show that the difference between the bootstrap distribution and the normal distribution is a suitable proxy for the difference between the exact distribution and the normal distribution, hence it is a reasonable indicator of the identification strength.

1.4.1 Model Setup

$$\begin{cases} Y = X\theta + U \\ X = Z\Pi + V \end{cases}$$

$Y = (Y_1, ..., Y_n)', X = (X_1, ..., X_n)'$ are $n \times 1$ vectors of endogenous observations, and $Z = (Z_1, ..., Z_n)'$ is the $n \times k$ matrix of instruments, $k \geq 1$. $U = (u_1, ..., u_n)', V = (v_1, ..., v_n)'$, where the error term $(u_i, v_i)', i = 1, ..., n$, is assumed to have mean zero, and to be i.i.d., homoscedastic with covariance matrix $\Sigma = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}$. The parameter of interest is θ, and $\hat{\theta}_n$ is the IV estimator of θ:

$$\hat{\theta}_n \quad = \quad (X'P_zX)^{-1}X'P_zY$$

It is central to derive the distribution of $\hat{\theta}_n$ for statistical inference. The exact finite sample distribution of $\hat{\theta}_n$, however, is unknown without making further distributional assumptions. Instead, two alternative methods are often used to approximate the exact distribution in econometric applications: the limiting normal distribution, and the bootstrap distribution.

Under the conventional asymptotic theory where the $k \times 1$ vector Π is modeled as non-zero and fixed, the IV estimator $\hat{\theta}_n$ is asymptotically normally distributed as the sample size n gets large. In contrast, to explore the distribution of $\hat{\theta}_n$ when the instruments are only weakly related to the endogenous variable, Staiger and Stock (1997) develop weak instrument asymptotics, i.e. Π is modeled as local to zero.

Assumption 2. *(a)* $\Pi = \Pi_0 \neq 0$, *and* Π_0 *is fixed; (a')* $\Pi = \Pi_n = \frac{C}{\sqrt{n}}$, *and* C *is fixed.*

The asymptotics under Assumption 2(a) are called *Strong Instrument Asymptotics*, and the asymptotics under Assumption 2(a') are called *Weak Instrument Asymptotics*. (a)(a') are two alternative rank conditions, and in (a') the rank condition is only weakly satisfied. The following results and notations are used: $Z'Z/n \xrightarrow{p} Q_{zz} \equiv E(Z_i'Z_i)$, $(\frac{Z'U}{\sqrt{n}}, \frac{Z'V}{\sqrt{n}}) \Rightarrow (\Psi_{zu}, \Psi_{zv})$, and $(\Psi_{zu}, \Psi_{zv})'$ is distributed $N(0, \Sigma \otimes Q_{zz})$. The validity of these results follows from law of large numbers and a central limit theorem, after assuming the existence of second moments. By the similar derivation as in Staiger and Stock (1997), the following two well-known results hold.

Under *Strong Instrument Asymptotics*:

$$\sqrt{n}(\hat{\theta}_n - \theta) \;\; \Rightarrow \;\; (\Pi_0' Q_{zz} \Pi_0)^{-1} \Pi_0' \Psi_{zu} \sim N(0, (\Pi_0' Q_{zz} \Pi_0)^{-1} \sigma_u^2) \tag{1.6}$$

Under *Weak Instrument Asymptotics*:

$$\hat{\theta}_n - \theta \;\; \Rightarrow \;\; [(Q_{zz}C + \Psi_{zv})' Q_{zz}^{-1} (Q_{zz}C + \Psi_{zv})]^{-1} (Q_{zz}C + \Psi_{zv})' Q_{zz}^{-1} \Psi_{zu} \tag{1.7}$$

If $k = 1$, i.e. the model is exactly identified, then (1.7) reduces to:

$$\hat{\theta}_n - \theta \;\; \Rightarrow \;\; (Q_{zz}C + \Psi_{zv})^{-1} \Psi_{zu}$$

The conventional result of (1.6) indicates that when instruments are strong, the IV estimator $\hat{\theta}_n$ is both consistent and asymptotically normally distributed. In contrast, the result of (1.7) indicates that the IV estimator $\hat{\theta}_n$ is neither consistent nor asymptotically normally distributed, if the rank condition is weak. As the magni-

tude of Π increases, however, the distribution of $\hat{\theta}_n$ in (1.7) gets closer to the normal distribution in (1.6). In the extreme case that $C = \sqrt{n}\Pi_0$, (1.6) and (1.7) coincide. Π, the vector of nuisance parameters, is thus the driving force of the linear IV regression model: it determines whether θ can be consistently estimated, and whether the estimator $\hat{\theta}_n$ can be well approximated by the normal distribution.

As a function of Π, the concentration parameter μ^2 is a unit-less measure of the identification strength in the studies of weak instruments:

$$\mu^2 = \frac{\Pi' Z' Z \Pi}{\sigma_v^2}$$

The greater μ^2, the stronger the identification of the parameter θ, and the distribution of $\hat{\theta}_n$ gets closer to the normal distribution, as shown in Rothenberg (1984). In addition, Stock and Yogo (2005) suggest that there is a threshold of the concentration parameter for the set of weak instruments, i.e. instruments as well as identification are considered weak if μ^2/k is under the threshold. The first stage F test is suggested in Stock and Yogo (2005) to check whether the threshold is exceeded: if the F statistic is greater than the tabled critical values, typically around 10 for small k, then instruments as well as identification are not weak:

$$F = \frac{\hat{\Pi}_n' Z' Z \hat{\Pi}_n / k}{\hat{\sigma}_v^2}$$

where $\hat{\Pi}_n = (Z'Z)^{-1} Z'X$, $\hat{\sigma}_v^2 = (X - Z\hat{\Pi}_n)'(X - Z\hat{\Pi}_n)/(n-k)$.

As an alternative to the limiting normal distribution, the bootstrap provides another way of approximating the distribution of $\hat{\theta}_n$. For the linear IV regression model under homoscedasticity, the residual bootstrap is a commonly used bootstrap method. See, for example, Moreira *et al.* (2009). This bootstrap procedure or algo-

rithm is described as follows.

1. $\hat{U}, \hat{V}$ are the residuals induced by $\hat{\theta}_n, \hat{\Pi}_n$ in the linear IV model:

$$\hat{U} = Y - X\hat{\theta}_n, \ \hat{V} = X - Z\hat{\Pi}_n$$

2. Re-center $\hat{U}, \hat{V}$ to yield $\tilde{U}, \tilde{V}$, by pre-multiplying a constant matrix M_e, where $M_e = I_n - P_e$, and e is the n by 1 vector of ones:

$$\tilde{U} = M_e\hat{U}, \ \tilde{V} = M_e\hat{V}$$

3. Sampling the rows of $(\tilde{U}, \tilde{V})$ and Z independently n times with replacement, and let (U^*, V^*) and Z^* denote the outcome. The dependent variables (X^*, Y^*) are generated by:

$$\begin{cases} Y^* = X^*\hat{\theta}_n + U^* \\ X^* = Z^*\hat{\Pi}_n + V^* \end{cases}$$

4. As the counterpart of the IV estimator $\hat{\theta}_n$, the bootstrapped IV estimator $\hat{\theta}^*_n$ is computed by the bootstrap resample (X^*, Y^*, Z^*):

$$\hat{\theta}^*_n \ = \ (X^{*'}P_{z^*}X^*)^{-1}X^{*'}P_{z^*}Y^*$$

5. Re-do Steps 2-4 B times, and $\{\hat{\theta}^{*i}_n, i = 1, ..., B\}$ are B i.i.d. estimators.

The bootstrap data generation process (D.G.P.) above aims to mimic the D.G.P. of the linear IV regression model: when instruments are strong, the equation $\hat{\Pi}_n = \Pi + O_p(n^{-1/2})$ indicates that $\hat{\Pi}_n$ is not substantially different from Π; in addition, the

variance of the bootstrap error term (u_i^*, v_i^*) converges to Σ, the variance of (u_i, v_i). Consequently, it is natural to expect that the mimicking process works well under strong instruments, and the distributions of $\hat{\theta}_n^*$ and $\hat{\theta}_n$ are alike. This conjecture on the bootstrapped estimator $\hat{\theta}_n^*$ is confirmed by the result below.

Under *Strong Instrument Asymptotics*:

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \quad \Rightarrow \quad (\Pi_0' Q_{zz} \Pi_0)^{-1} \Pi_0' \Psi_{zu} \tag{1.8}$$

The result of (1.8) motivates the usage of the bootstrap as a tool to detect the identification strength: under strong identification, the distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ is asymptotically identical to the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, and the asymptotic distribution is normal; if the distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ is found to be substantially different from normal, then it indicates that identification is weak.

The bootstrap mimicking process also helps explain why the bootstrap becomes problematic when instruments are weak: first of all, if Π is local to zero, then the connection $\hat{\Pi}_n = \Pi + O_p(n^{-1/2})$ implies that the difference between $\hat{\Pi}_n$ and Π becomes substantial, hence the identification strength in the bootstrap resample is substantially different from the identification strength in the sample; secondly, when $\hat{\theta}_n$ does not consistently estimate θ under weak instruments, the residual $\hat{U}$ does not converge to the error term U, since $\hat{U} = U - X(\hat{\theta}_n - \theta)$. Both of these two facts contribute to that the bootstrap D.G.P. does not mimic the D.G.P. of the linear IV model well under weak instruments, and consequently, the distributions of $\hat{\theta}_n^*$ and $\hat{\theta}_n$ are different.

To compare the identification strength in the bootstrap resample (X^*, Y^*, Z^*) with the identification strength in the sample (X, Y, Z), consider the concentration parameter μ^{2*}, the bootstrap counterpart of μ^2:

$$\mu^{2*} = \frac{\hat{\Pi}_n' Z^{*'} Z^* \hat{\Pi}_n}{\sigma_v^{2*}}, \text{where } \sigma_v^{2*} = \frac{V^{*'} V^*}{n}$$

Under *Strong Instrument Asymptotics*:

$$\mu^2 \to \infty, \text{and } \mu^{2*} \to \infty \tag{1.9}$$

Under *Weak Instrument Asymptotics*:

$$\mu^2 \xrightarrow{p} \frac{C' Q_{zz} C}{\sigma_v^2}, \text{and } \mu^{2*} \Rightarrow \frac{C' Q_{zz} C + 2C'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1} \Psi_{zv}}{\sigma_v^2} \tag{1.10}$$

The result of (1.9) states that when identification is strong, μ^2 and μ^{2*} go to infinity, implying that the distributions of $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are both asymptotically normal, as stated in (1.6) and (1.8). On the contrary, the result of (1.10) states that when identification is weak, μ^{2*} and μ^2 do not go to infinity, and are asymptotically different, which implies that the distributions of $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are both asymptotically non-normal, and their asymptotical distributions are not identical. The asymptotic difference in μ^2 and μ^{2*} is $\frac{2C'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1} \Psi_{zv}}{\sigma_v^2}$: under homoscedasticity, it has mean k.

Table 1.1 reports the difference between μ^2 and μ^{2*} when $k = 1$ by Monte Carlo studies: the relative difference is found to be substantial when μ^2 is small, and negligible when μ^2 is large; overall, μ^{2*} is greater than μ^2. Another interpretation of (1.10) is that, loosely speaking, the F statistic of the bootstrap resample is above the

F statistic of the sample by 1, because of the relation $E(F) \approx \mu^2/k+1$ in Stock *et al.* (2002).

The comparison of μ^{2*} and μ^2 indicates a useful result in the linear IV model where μ^2/k and μ^{2*}/k are measures of the identification strength: on average, the identification strength in the bootstrap resample is similar to, and slightly stronger (plus 1) than the identification strength in the sample; consequently, if the identification strength in the bootstrap resample is weak, then the identification strength in the sample must also be weak. In this sense, the proposed bootstrap strategy for detecting weak identification is conservative.

The Edgeworth expansion provides another look at the bootstrap strategy for detecting weak identification. For the purpose of this chapter, it suffices to consider the two-term expansion of the standardized estimator $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$, where $\sigma^2 = (\Pi_0' Q_{zz} \Pi_0)^{-1}\sigma_u^2$.

Define $R_i = (Z_i X_i, Z_i Y_i, vec(Z_i' Z_i)')'$, $\mu = E(V_i) = (Q_{zz}\Pi, Q_{zz}\Pi\beta, vec(Q_{zz})')'$. Rewrite $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ in the form of $\sqrt{n}A(\bar{R})$, where $\bar{R} = \frac{1}{n}\sum_{i=1}^n R_i$, and $A(\mu) = 0$. The following result is the application of the smooth function model and Theorem 2.2 in Hall (1992). A similar result is available in Moreira *et al.* (2009).

Theorem. *Under Assumption 2(a), and assume two conditions: (i)* $E(||R_i||^3) < \infty$, *(ii) lim* $sup_{||t||\to\infty}|Eexp(it'R_i)| < 1$, $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ *and its bootstrap counterpart admit two-term Edgeworth expansions uniformly in* x*:*

$$P(\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \leq x) = \Phi(x) + n^{-1/2}p(x)\phi(x) + o(n^{-1/2}) \tag{1.11}$$

$$P(\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) = \Phi(x) + n^{-1/2}p^*(x)\phi(x) + o(n^{-1/2}) \tag{1.12}$$

where $p(x)$ *is a polynomial of degree 2, with coefficients depending on* θ, Π*, and moments of* R_i *up to order 3,* $p^*(x)$ *is the bootstrap counterpart of* $p(x)$ *with coefficients*

depending on $\hat{\theta}_n, \hat{\Pi}_n$*, and moments of* R_i^**.*

Figure 1.2 is drawn based on the results of the Edgeworth expansion. The distance between the exact distribution of $\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}}$ and the standard normal distribution has order $O(n^{-1/2})$, which is the same as the order of the distance between the bootstrap distribution of $\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ and the standard normal distribution. Compared with the normal distribution, the bootstrap distribution has the well known property of asymptotic refinement: it is closer to the the exact distribution, as the distance has order $O(n^{-1})$, which results from $p^*(x) - p(x) = O(n^{-1/2})$: $\hat{\theta}_n, \hat{\Pi}_n$, and moments of R_i^* approach θ, Π, and moments of R_i at rate $n^{-1/2}$.

As discussed above, the bootstrap strategy for detecting weak identification is to use the distance between the bootstrap distribution and the normal as a proxy of the distance between the exact distribution and the normal distribution. This strategy is supported by the Edgeworth expansion in two ways: firstly, these two distances have the same order $O(n^{-1/2})$; secondly, the price of using the proxy is low, i.e. the proxy error is $O(n^{-1})$, because $[P(\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \leq x) - \Phi(x)] - [P(\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) - \Phi(x)] = O(n^{-1})$.

To summarize, the bootstrap distribution is a good proxy for the exact distribution for the purpose of this chapter. Under the null hypothesis of strong identification, the disparity between the exact distribution and the normal distribution is reflected by the disparity between the bootstrap distribution and the normal distribution. If the bootstrap distribution is found to be substantially different from normal, then it is appropriate to conclude that the exact distribution is substantially different from normal as well, hence the identification strength is weak.

So far, the discussion is restricted to the linear IV model under the homoscedasticity assumption. However, it is well understood that this assumption is unlikely

to hold in practice: for example, $E(u_i^2|Z_i)$ may not be constant, but depend on Z_i. As a result, it is common that empirical researchers need to take the existence of heteroscedasticity into consideration.

Once the homoscedasticity assumption is loosened, the validity of the popular F test for detecting weak instruments/identification is under doubt: the way in which Stock and Yogo (2005) derive the critical values of the F test crucially depends on the homoscedasticity assumption. Consequently, it is not clear whether these critical values can still be used when heteroscedasticity instead of homoscedasticity takes place.

The validity of the bootstrap test approach, on the contrary, stays unaffected, when the pair bootstrap replaces the residual bootstrap. Freedman (1984) shows that the pair bootstrap of the IV estimator remains valid under heteroscedasticity: the IV estimator and its bootstrap counterpart asymptotically have the same normal distribution under strong instruments. The idea of the pair bootstrap is to directly resample the data. In the special case of $k = 1$, the pair bootstrap is to draw the bootstrap resample (X^*, Y^*, Z^*) from the empirical distribution of (X, Y, Z). Compared with the residual bootstrap above, the pair bootstrap is non-parametric, and it preserves the possible heteroscedastic relations in the IV model, as proved in Freedman (1984). Consequently, if heteroscedasticity is concerned, the bootstrap test procedure involves two minor modifications: (i) use the pair bootstrap to compute $\hat{\theta}_n^*$; (ii) use a heteroscedasticity-robust estimator for $\hat{\sigma}$.

1.4.2 An Empirical Example: Card (1995)

In this section, an empirical application is investigated to illustrate the bootstrap approach for detecting weak identification. The same data as in Card (1995) is used. By employing the IV approach, Card (1995) answers the following question: what is the return to education? Or specifically, how much more can an individual earn if he/she completes an extra year of schooling?

The dataset is ultimately taken from the National Longitudinal Survey of Young Men between 1966-1981 with 3010 observations, and there are two variables in the dataset that measure college proximity: $nearc2$ and $nearc4$, both are dummy variables, and are 1 if there is a 2-year, 4-year college in the local area respectively. See Card (1995) for the detailed description of the data. To identify the return to education, Card (1995) considers a structural wage equation as follows:

$$lwage = \alpha + \theta edu + W'\beta + u$$

where $lwage$ is the log of wage, edu is the years of schooling, the covariate vector W contains the control variables, and u is the error term. Among the set of parameters (α, θ, β'), θ measuring the return to education is of interest.

In the basic specification, Card (1995) uses five control variables: experience, the square of experience, the dummy for race, the dummy for living in the south, and the dummy for living in the standard metropolitan statistical area (SMSA). To bypass the issue that experience is also endogenous, Davidson and MacKinnon (2010) replace experience, the square of experience with age, the square of age. Following Davidson and MacKinnon (2010), I used age, square of age, and the three dummy variables as control variables, hence edu is the only endogenous regressor. In

addition, an extra instrument $nearc2 \wedge nearc4$ is constructed: $nearc2 \wedge nearc4$ is a dummy variable, and is 1 if there are both a 2-year and a 4-year college in the local area. Unlike Davidson and MacKinnon (2010), I use the three instruments, $nearc2$, $nearc2 \wedge nearc4$, $nearc4$, one by one as the single instrument for edu to better illustrate the approach discussed in this chapter, while Davidson and MacKinnon (2010) simultaneously use more than one instrument.

The identification strength under the three potential IV's is examined by the first stage F test in Stock and Yogo (2005): if $nearc2$ is used as the IV, $F \approx 0.54$; if $nearc2 \wedge nearc4$ is used as the IV, $F \approx 6.98$; if $nearc4$ is used as the IV, $F \approx 10.22$. According to the rule of thumb $F > 10$ suggested in Stock and Yogo (2005), these F statistics suggest that $nearc4$ is a strong IV, $nearc2$, $nearc2 \wedge nearc4$ are not. Table 1.2 reports the point estimate and 95% confidence interval of θ under each of these instruments. If the point estimate of the return to education is of interest, 0.0936 derived by $nearc4$ is more reliable, compared with the other point estimates. Based on these empirical results, an additional year of education increases the wage by about 9.36%; however, the possibility that this effect is zero can not be rejected at 95%.

As proposed in this chapter, the bootstrap can help evaluate the identification strength. To allow for heteroscedasticity, the pair bootstrap is employed, and the IV estimator of θ is computed $B = 9999$ times, using the three instruments one by one: specifically, the bootstrap resample is directly drawn from the sample with replacement, and the size of the resample is equal to the number of observations; the bootstrapped IV estimator is computed by each bootstrap resample, and this process is replicated B times. To make it comparable to the standard normal variate, the bootstrapped estimator is standardized by subtracting the IV estimate and dividing by the standard error of the IV estimator, where the standard error is computed in

the way of White (1980). The p.d.f and quantiles of the bootstrapped IV estimator after standardization are plotted against the standard normal variate in Figure 1.3. Figure 1.3 shows that, the bootstrap distribution is closer to the normal distribution, when the instrument is stronger. The figure of the bootstrap distribution and the Q-Q plot[7] hence are useful tools to help detect weak identification, since they provide empirical researchers a graphic evaluation of the identification strength.

Table 1.2 reports the 95% C.I. of the return to education by the bootstrap percentile method, in addition, the identification-robust conditional likelihood ratio (CLR)[8] test by Moreira (2003) is also applied to construct a C.I. for comparison. Instead of the F test, the proposed b test is applied to determine whether the strength of identification is strong or weak, with the tentative threshold $\gamma = 0.25$.

*nearc*2: the bootstrap C.I. of the return to education is $(-4.3279, 4.8868)$, while the C.I. by t/Wald is $(-0.8188, 1.8346)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 2.46 > \gamma$, $b_1 \approx 12.09 > z_{95\%}$, hence the null of strong identification under *nearc*2 is rejected.

*nearc*2 $\wedge$ *nearc*4: the bootstrap C.I. of the return to education is found to be $(0.0019, 0.4664)$, while the C.I. by t/Wald is $(-0.0099, 0.2692)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 0.66 > \gamma$, $b_1 \approx 6.36 > z_{95\%}$, hence the null of strong identification under *nearc*2 $\wedge$ *nearc*4 is rejected.

*nearc*4: the bootstrap C.I. of the return to education is found to be $(0.0034, 0.2579)$, while the C.I. by t/Wald is $(-0.0027, 0.1899)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 0.32 > \gamma$, $b_1 \approx 2.71 > z_{95\%}$, hence the null of

[7]Compared to the p.d.f., the Q-Q plot is known to be a better way of comparing distributions.

[8]Note: The IV model under consideration is just identified, hence CLR is equivalent to the AR test in Anderson and Rubin (1949) and K test in Kleibergen (2002).

strong identification under $nearc4$ is rejected.

To conclude, both the F test and the bootstrap based b test can detect the weak instrument $nearc2$, $nearc2 \wedge nearc4$ and support the view that $nearc4$ is the strongest among the three potential instruments. The difference is, the b test considers $nearc4$ as weak, while the F test treats $nearc4$ as strong (but just across the threshold). It is surprising that, although $F > 10$ holds under $nearc4$, the relative difference of the C.I. by t/Wald and the bootstrap C.I. is as large as 0.32. It thus indicates that the $F > 10$ rule is not strict enough, i.e. the disparity between the two C.I.'s is still severe, although $F > 10$ holds in this example.

The b and F tests are proposed on different grounds, i.e. the b test is based on the comparison of the length of confidence intervals, while the F test is based on the threshold of the rank condition, hence it is not surprising that these two tests can imply different outcomes. In contrast with the F test, the b test has its advantages: firstly of all, it provides a graphic view of the identification strength, i.e. the Q-Q plot in Figure 1.3 shows that the bootstrap distribution is not very close to the normal distribution, hence the identification strength appears weak; secondly, unlike the F test, the b test does not rely on the restrictive homoscedasticity assumption, and has the potential of being extended to the generalized GMM framework.

1.5 Simulation

This section presents Monte Carlo results to evaluate the power of the proposed b test. The disparity between the two intervals, (1.3) and (1.4), is reported. The threshold γ is calibrated to the $F > 10$ rule.

The linear IV model described in Section 1.4.1 is employed in the D.G.P., with the following choice of parameters: $\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, where $\rho \in \{0.99, 0.50, 0.01\}$ to introduce high, moderate and low degrees of endogeneity, respectively; $\theta = 0$; $z_i \sim NID(0,1)$, $i = 1, ..., n$, and $n = 1000$.

A sequence of μ^2, $\mu^2 \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 40, 60, 80\}$, is chosen by assigning different values to Π. For each μ^2, the data of (X, Y, Z) is generated by the linear IV model with the parameters specified above, and the bootstrapped IV estimator $\hat{\theta}_n^*$ is computed $B = 9999$ times by the residual bootstrap. The number of replications equals 1000.

The results of the Monte Carlo studies are reported in Table 1.3.

C_0 **vs.** C_t: C_0 denotes the interval derived by taking the 2.5%, 97.5% quantiles of $\hat{\theta}_n$. C_t is the 95% C.I. derived by inverting the t test. Table 1.3 reports the relative difference in lengths of these two intervals: the median absolute difference in the lengths of C_0, C_t weighted by the length of C_t is reported. The departure of C_t from C_0 is more severe for larger ρ, and as expected, the departure of C_t from C_0 shrinks as μ^2 increases. It is found that the relative difference can be as high as 0.62 when $\mu^2 = 10$. This is a bit surprising since the identification strength under $\mu^2 = 10$ is generally not considered as very weak: as reported in the table, there is about half chance that the F test will consider the identification under $\mu^2 = 10$ as *strong*. In other words, although the F test and its decision rule may treat the identification strength under $\mu^2 = 10$ as *strong*, the departure of C_t from C_0 can still be severe.

C_b **vs.** C_t: the 95% C.I. C_b by the bootstrap is compared with C_t. Table 1.3 reports the median absolute value of the relative difference in the lengths of C_b, C_t.

As expected, the difference shrinks as μ^2 increases. In particular, when $\mu^2 > 10$, the relative difference does not exceed 0.50 in absolute value; when $\mu^2 \geq 20$, the relative difference does not exceed 0.25 in absolute value; when $\mu^2 \geq 80$, the difference is negligible.

b **test**: the proposed b test is applied to test the null of strong identification, and the percentage of rejecting strong identification is reported. Two tentative cutoffs, $\gamma = 0.25$, 0.50 are considered. Under the stricter rule of $\gamma = 0.25$, the b test rejects the null more often.

F **test**: the F test and its decision rule in Stock and Yogo (2005) are applied to provide a benchmark for the b test. Table 1.3 reports the percentage of concluding weak identification by the F test (the frequency of the F statistic is less than its critical value) for different ρ's. As discussed above, the F test examines whether μ^2/k exceeds the cutoff, hence it does not depend on ρ, the degree of endogeneity.

By comparing the performance of the b test with the F test, $\gamma = 0.25$ appears to be a reasonable choice. It corresponds to μ^2 around 10 with endogeneity close to zero. With this threshold, the frequency of concluding weak identification by the b test is comparable to the F test, although b rejects strong identification slightly more often. This chapter hence suggests $\gamma = 0.25$ as the quantitative threshold for distinguishing strong and weak identification, and this simple rule is not off the mark for practical reasons.

1.6 Conclusion

This chapter suggests that the bootstrap is a useful tool for detecting weak identification in IV/GMM applications. The distinguishing feature of the bootstrap based approach is that it provides a graphic view of the identification strength. By eyeballing the graph of the bootstrap distribution, and comparing it with the normal distribution through the Q-Q plot, empirical researchers can evaluate whether or not weak identification exists. The underlying reason is simple: strong identification implies the bootstrap distribution is close to, and asymptotically identical to the normal distribution.

A quantitative threshold for distinguishing *strong* and *weak* identification is suggested based on the comparison of two C.I.'s for the parameter of interest: the C.I. by inverting the t/Wald test and the bootstrap percentile C.I.. The difference of these two C.I.'s boils down to the difference between the bootstrap distribution and the normal distribution, and exceeding the threshold implies that the relative difference of the two C.I.'s is at least as large as a quarter. For practical purposes, the identification strength is considered as *weak* in this chapter once this threshold is exceeded. Monte Carlo experiments show that this threshold is comparable to and slightly stricter than the $F > 10$ rule of thumb in Stock and Yogo (2005). Even in the i.i.d. and homoscedasticity setting, $F > 10$ is found to be not strict enough, when it comes to the comparison of C.I.'s: the relative difference of the two commonly used C.I.'s named above can be very large, even when $F > 10$ holds; an application to Card (1995) also makes the same point.

Figure 1.1: Three Related Distributions

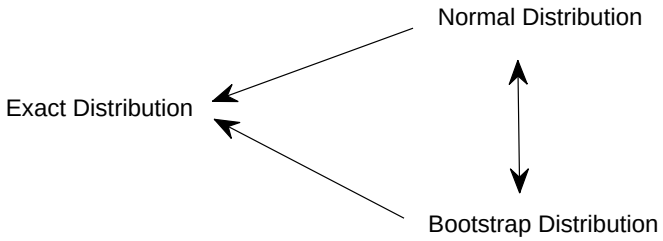
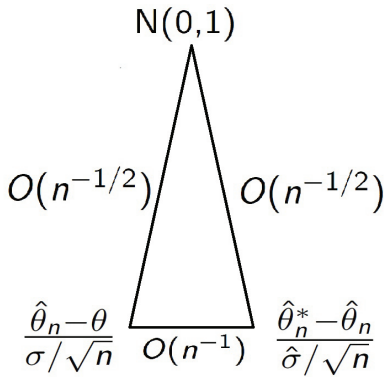


Figure 1.2: Expressions and Distances



Notes: The normal distribution and the bootstrap distribution are two approximations to the exact distribution of IV/GMM estimators. If the exact distribution of $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ is well approximated by the normal distribution $N(0,1)$, then the bootstrap distribution of $\frac{\hat{\theta}^*_n-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ is well approximated by $N(0,1)$ as well, because of the same magnitude $O(n^{-1/2})$ of the approximation error. A substantial difference between $\frac{\hat{\theta}^*_n-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ and $N(0,1)$ indicates the difference between $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ and $N(0,1)$ is also substantial, hence provides the evidence of weak identification.

Table 1.1: A Monte Carlo study of μ^2 and μ^{2*}

	μ^2												
	2	4	6	8	10	12	14	16	18	20	40	60	80
μ^{2*}	3.1	5.1	7.1	9.1	11.2	13.2	15.2	17.2	19.2	21.2	41.4	61.5	81.6

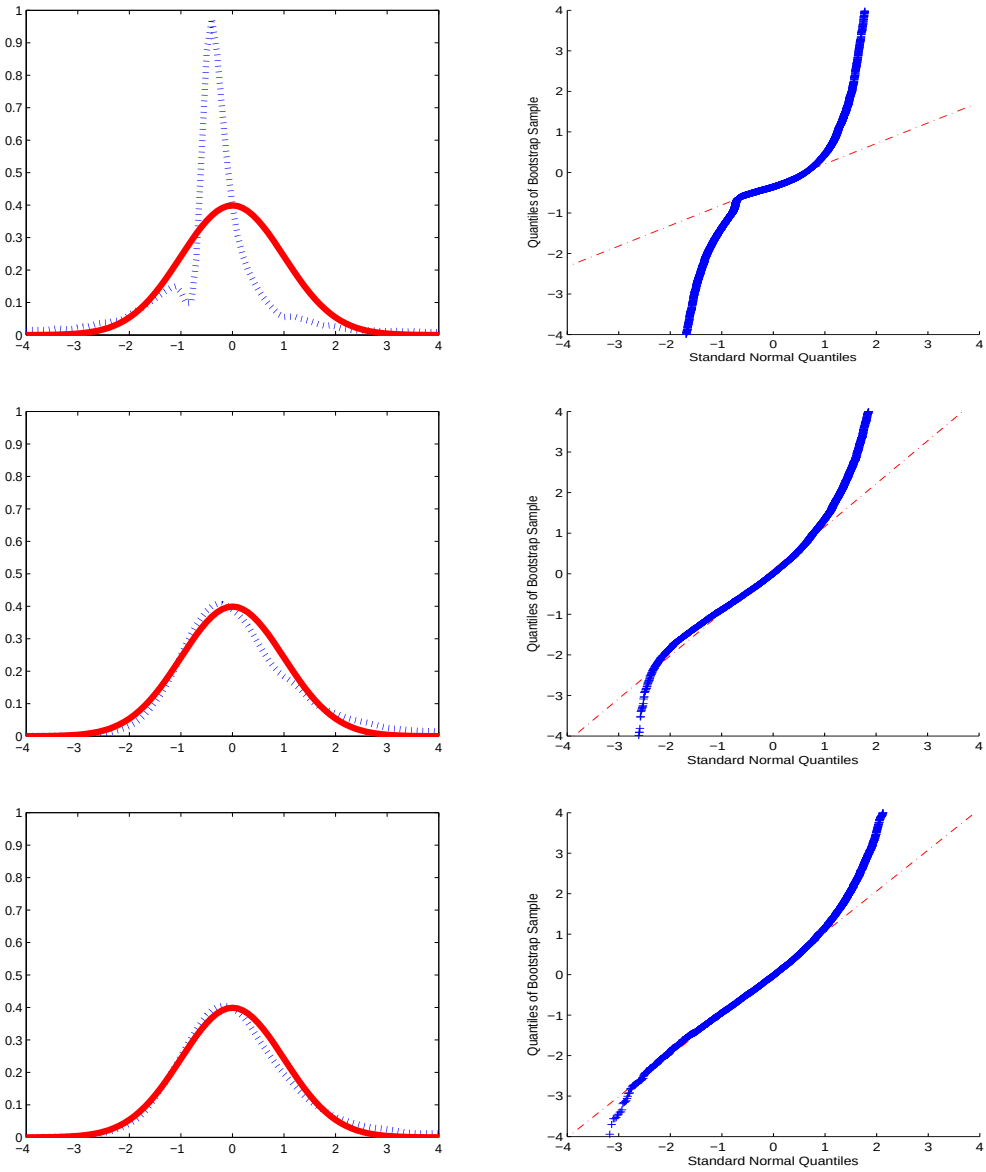
Notes: This table compares the concentration parameter μ^2 with μ^{2*}, the bootstrap counterpart of μ^2, by a Monte Carlo study. For each μ^2, the data of X, Z, V are generated by $x_i = z_i\Pi + v_i$, where: (i) x_i, z_i, v_i are the i^{th} elements of X, Z, V; (ii) $z_i \sim NID(0,1)$, $v_i \sim NID(0,1)$, $i = 1, ..., 1000$; (iii) Π is determined by the value of μ^2. The reported μ^{2*} is the sample average of 1000 replications; in each replication, the residual bootstrap is conducted 1000 times.

Table 1.2: Return to Education

IV	$nearc2$	$nearc2 \wedge nearc4$	$nearc4$
F statistic	0.54	6.98	10.22
$\hat{\theta}_n$	0.5079	0.1297	0.0936
95% C.I. by t/Wald	(-0.8188, 1.8346)	(-0.0099, 0.2692)	(-0.0027, 0.1899)
by bootstrap	(-4.3279, 4.8868)	(0.0019, 0.4664)	(0.0034, 0.2579)
by AR/CLR/K	$(-\infty, -0.1750] \cup [0.0867, +\infty)$	$[0.0133, 0.5253]$	[0.0009, 0.2550]

Notes: This table presents the estimate $\hat{\theta}_n$ and confidence interval for return to education using the data of Card (1995). The first stage F statistic is reported for the three instrumental variables, $nearc2$, $nearc2 \wedge nearc4$, $nearc4$, which are used one by one for the endogenous years of schooling. The included control variables are age, square of age, and three dummy variables for race, living in the south, living in the SMSA.

Figure 1.3: The bootstrap distribution, p.d.f and Q-Q plot



Notes: The p.d.f. and Q-Q plot of the bootstrap distribution are presented, under three instrumental variables in the application of Card (1995). 1^{st} **row:** $nearc2$ as IV; 2^{nd} **row:** $nearc2 \wedge nearc4$ as IV; 3^{rd} **row:** $nearc4$ as IV; **Left:** p.d.f of the bootstrapped IV estimator after standardization (dotted) against the standard normal (solid); **Right:** the Q-Q plot. 9999 bootstrap replications are conducted.

Table 1.3: The performance of b, F for detecting weak identification

	μ^2												
	2	4	6	8	10	12	14	16	18	20	40	60	80
Comparison of C.I.													
C_0 vs. C_t:													
$\rho = 0.99$	5.24	1.64	1.37	0.80	0.62	0.60	0.50	0.41	0.38	0.37	0.22	0.18	0.14
0.50	2.62	1.10	0.83	0.62	0.41	0.38	0.33	0.29	0.29	0.26	0.15	0.13	0.11
0.01	1.90	0.95	0.53	0.39	0.32	0.27	0.23	0.21	0.21	0.19	0.12	0.10	0.08
C_b vs. C_t:													
$\rho = 0.99$	1.87	1.61	1.10	0.86	0.61	0.45	0.37	0.33	0.27	0.23	0.11	0.07	0.05
0.50	1.56	1.06	0.72	0.47	0.37	0.29	0.22	0.19	0.16	0.14	0.07	0.05	0.03
0.01	1.45	0.97	0.61	0.40	0.29	0.22	0.18	0.15	0.13	0.12	0.05	0.04	0.03
Rejection freq. of b													
$\gamma = 0.50$:													
$\rho = 0.99$	90.1	84.3	75.9	65.4	53.9	40.3	30.4	22.9	14.4	10.1	0.1	0.0	0.0
0.50	83.3	69.8	57.7	43.5	34.0	24.3	16.9	13.6	9.0	5.7	0.0	0.0	0.0
0.01	81.4	68.2	52.3	38.3	27.4	18.2	12.3	7.7	4.7	3.0	0.0	0.0	0.0
$\gamma = 0.25$:													
$\rho = 0.99$	96.4	96.0	94.4	89.7	81.7	73.3	63.1	56.6	44.7	35.5	1.6	0.0	0.0
0.50	93.6	87.5	82.6	67.6	60.3	49.5	37.9	29.0	23.8	17.9	0.6	0.0	0.0
0.01	92.2	87.1	77.2	62.2	50.2	39.2	29.5	22.6	15.7	11.5	0.2	0.0	0.0
Rejection freq. of F													
$\rho = 0.99$	94.1	82.7	70.5	53.0	43.3	32.1	22.6	16.1	12.5	7.9	0.0	0.0	0.0
0.50	93.9	82.1	66.3	56.9	42.9	32.7	22.4	15.7	12.5	8.2	0.0	0.0	0.0
0.01	93.4	83.7	69.4	54.9	45.7	33.3	22.3	15.2	11.6	7.4	0.0	0.0	0.0

Notes: This table presents the Monte Carlo results of comparing the bootstrap based b test proposed in this chapter with the F test of Stock and Yogo (2005). The frequencies of concluding weak identification by the b test is reported for each μ^2, and the frequencies of concluding weak identification by the F test is also reported. μ^2 is the concentration parameter, and the greater μ^2 is, the stronger the strength of identification; ρ is the correlation coefficient, and the greater ρ is, the stronger the endogeneity; C_0, C_t and C_b are the C.I. derived by taking the 2.5%, 97.5% quantiles of $\hat{\theta}_n$, the 95% C.I. by inverting t, the 95% C.I. by the bootstrap respectively, and their relative difference in length (the median absolute value) is reported; γ is the threshold, identification is considered as *weak* by the b test if the relative difference between C_t and C_b exceeds γ.

Appendix

Proof. (1.5)

By the joint normal distribution of quantile estimators, the difference of two quantile estimators is also normally distributed:

$$\sqrt{B}\left[\left(\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}-\frac{\hat{\theta}^*_{n,\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}\right)-(q_{1-\alpha/2}-q_{\alpha/2})\right]\Rightarrow N(0,\Omega_{11}+\Omega_{22}-2\Omega_{12})$$

Rewrite the LHS:

$$\sqrt{B}\left[\left(\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}}-1\right)-\left(\frac{q_{1-\alpha/2}-q_{\alpha/2}}{2z_{1-\alpha/2}}-1\right)\right]\cdot 2z_{1-\alpha/2}\Rightarrow N(0,\Omega_{11}+\Omega_{22}-2\Omega_{12})$$

The result follows after the substitution of $\hat{D}=\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}}-1$, $D=\frac{q_{1-\alpha/2}-q_{\alpha/2}}{2z_{1-\alpha/2}}-1$.

□

Proof. (1.6)

$$\begin{aligned}
\frac{X'P_zX}{n} &= \frac{X'Z(Z'Z)^{-1}Z'X}{n}\\
&= X'Z(Z'Z)^{-1}\frac{Z'Z}{n}(Z'Z)^{-1}Z'X\\
&\xrightarrow{p} \Pi_0'Q_{zz}\Pi_0\\
\frac{X'P_zU}{\sqrt{n}} &= X'Z(Z'Z)^{-1}\frac{Z'U}{\sqrt{n}}\\
&\Rightarrow \Pi_0'\Psi_{zu}\\
\sqrt{n}(\hat{\theta}_n-\theta) &= (\frac{X'P_zX}{n})^{-1}\frac{X'P_zU}{\sqrt{n}}\\
&\Rightarrow (\Pi_0'Q_{zz}\Pi_0)^{-1}\Pi_0'\Psi_{zu}
\end{aligned}$$

□

Proof. (1.7)

$$\begin{aligned} X'P_zX &= \frac{X'Z}{\sqrt{n}}(\frac{Z'Z}{n})^{-1}\frac{Z'X}{\sqrt{n}} \\ &= (\frac{Z'Z}{n}C + \frac{Z'V}{\sqrt{n}})'(\frac{Z'Z}{n})^{-1}(\frac{Z'Z}{n}C + \frac{Z'V}{\sqrt{n}}) \\ &\Rightarrow (Q_{zz}C + \Psi_{zv})'Q_{zz}^{-1}(Q_{zz}C + \Psi_{zv}) \end{aligned}$$

$$\begin{aligned} X'P_zU &= \frac{X'Z}{\sqrt{n}}(\frac{Z'Z}{n})^{-1}\frac{Z'U}{\sqrt{n}} \\ &\Rightarrow (Q_{zz}C + \Psi_{zv})'Q_{zz}^{-1}\Psi_{zu} \\ \hat{\theta}_n - \theta &= (X'P_zX)^{-1}X'P_zU \\ &\Rightarrow [(Q_{zz}C + \Psi_{zv})'Q_{zz}^{-1}(Q_{zz}C + \Psi_{zv})]^{-1}(Q_{zz}C + \Psi_{zv})'Q_{zz}^{-1}\Psi_{zu} \end{aligned}$$

If exactly identified:

$$\begin{aligned} \hat{\theta}_n - \theta &= (Z'X)^{-1}Z'U \\ &\Rightarrow (Q_{zz}C + \Psi_{zv})^{-1}\Psi_{zu} \end{aligned}$$

□

Proof. (1.8)

$$\begin{aligned} \hat{\Pi}_n^* &\equiv (Z^{*'}Z^*)^{-1}Z^{*'}X^* \\ &= \hat{\Pi}_n + (\frac{Z^{*'}Z^*}{n})^{-1}\frac{Z^{*'}V^*}{n} \\ &= \Pi_0 + (\frac{Z^{'}Z}{n})^{-1}\frac{Z^{'}V}{n} + (\frac{Z^{*'}Z^*}{n})^{-1}\frac{Z^{*'}V^*}{n} \\ &\xrightarrow{p} \Pi_0 \\ \frac{Z^{*'}U^*}{\sqrt{n}} &\Rightarrow \Psi_{zu} \text{ by Lyapunov's Central Limit Theorem:} \end{aligned}$$

Let $m_i = Z_i^{*'} u_i^*$, rewrite $\frac{Z^{*'}U^*}{\sqrt{n}} = \frac{\sum_{i=1}^n Z_i^{*'} u_i^*}{\sqrt{n}} = \frac{\sum_{i=1}^n m_i}{\sqrt{n}}$. By construction, $m_i's$ are independent with mean $\mu_i = 0$, variance $\sigma_i^2 = \frac{Z'Z}{n}\frac{\tilde{U}'\tilde{U}}{n}$, and $\sigma_i^2 \xrightarrow{p} Q_{zz}\sigma_u^2$ under *Strong Instrument Asymptotics*. To verify the Lyapunov's condition: for $1 > \delta > 0$, the expected values $\mathbb{E}[|m_i|^{2+\delta}] < \infty$, and $\lim_{n\to\infty} \frac{1}{(\sum_{i=1}^n \sigma_i^2)^{\frac{2+\delta}{2}}} \sum_{i=1}^n \mathbb{E}[|m_i - \mu_i|^{2+\delta}] = \lim_{n\to\infty} \frac{O_p(n)}{O_p(n+\frac{n\delta}{2})} = 0$.

Combining the two results above:

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) &= \left[X^{*'}Z^*(Z^{*'}Z^*)^{-1}\frac{Z^{*'}Z^*}{n}(Z^{*'}Z^*)^{-1}Z^{*'}X^*\right]^{-1} X^{*'}Z^*(Z^{*'}Z^*)^{-1}\frac{Z^{*'}U^*}{\sqrt{n}} \\
&= \left[\hat{\Pi}_n^{*'}\left(\frac{Z^{*'}Z^*}{n}\right)\hat{\Pi}_n^*\right]^{-1}\hat{\Pi}_n^*\frac{Z^{*'}U^*}{\sqrt{n}} \\
&\Rightarrow (\Pi_0^{'}Q_{zz}\Pi_0)^{-1}\Pi_0^{'}\Psi_{zu}
\end{aligned}$$

□

Proof. (1.9)

$$\begin{aligned} \mu^2 &= \frac{\Pi' Z' Z \Pi}{\sigma_v^2} \\ &= O_p(n) \\ &\to \infty \end{aligned}$$

$$\hat{\Pi}_n \xrightarrow{p} \Pi,\ \sigma_v^{2*} \xrightarrow{p} \sigma_v^2,\ \frac{Z^{*'} Z^*}{n} \xrightarrow{p} Q_{zz}$$

$$\begin{aligned} \mu^{2*} &= \frac{\hat{\Pi}_n' Z^{*'} Z^* \hat{\Pi}_n}{\sigma_v^{2*}} \\ &= O_p(n) \\ &\to \infty \end{aligned}$$

□

Proof. (1.10)

$$\begin{aligned} \mu^2 &= \frac{C' \frac{Z'Z}{n} C}{\sigma_v^2} \\ &\xrightarrow{p} \frac{C' Q_{zz} C}{\sigma_v^2} \\ \mu^{2*} &= \frac{[C + (\frac{Z'Z}{n})^{-1} \frac{Z'V}{\sqrt{n}}]' \frac{Z^{*'} Z^*}{n} [C + (\frac{Z'Z}{n})^{-1} \frac{Z'V}{\sqrt{n}}]}{\sigma_v^{2*}} \\ &\Rightarrow \frac{C' Q_{zz} C + 2C' \Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1} \Psi_{zv}}{\sigma_v^2} \end{aligned}$$

□

CHAPTER **Two**

Weak Identification in (C)CAPM

2.1 Introduction

The Fama-MacBeth (FM) procedure analyzed by Shanken (1992) is widely used in empirical studies of the (C)CAPM. This two-pass procedure involves a first stage time series regression to derive β, the correlation matrix of financial assets with risk factors, and a second stage cross-sectional regression using β estimated from the first stage as regressors. The R^2 at the second stage is typically reported by empirical researchers to illustrate how much cross-sectional variation of asset returns can be explained by their proposed factors, and it often increases dramatically when a new factor is added: see, e.g., Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Acharya and Pedersen (2005), Lustig and Van Nieuwerburgh (2005), Li *et al.* (2006), Santos and Veronesi (2006), Hansen *et al.* (2008). Although it is common to evaluate how well models fit real data by looking at the R^2, what is surprising is that the empirical (C)CAPM literature heavily depends on the usage of the R^2 for model comparison. The various versions of the (C)CAPM with new risk factors are favored, at least partially due to their large R^2.

This chapter shows that the R^2 of the FM two-pass procedure can be large even when risk factors are irrelevant, hence a large R^2 does not imply risk factors are relevant. The underlying reason is, when factors are irrelevant, the β matrix at the first stage does not have full rank, which further induces a spurious distribution of the R^2 at the second stage. I call this problem *weak identification*, because it displays the similar feature as the weak instrument problem surveyed by Stock *et al.* (2002): the validity of the FM procedure crucially depends on the quality of regressors generated from its first stage; if irrelevant factors are included, the rank condition of the second stage is violated, hence the outcome of the FM procedure is no longer reliable.

I start by taking the influential work by Lettau and Ludvigson (2001) for example to illustrate the main point of this chapter. Lettau and Ludvigson (2001) report larger values of R^2 after adding a conditioning variable, the log consumption-wealth ratio *cay*, as a new factor. I consider four specifications of the (C)CAPM in Lettau and Ludvigson (2001), and replace *cay* with cay_{sim}, where cay_{sim} is randomly drawn from a normal distribution; the other settings remain unchanged. Table 2.1 presents the R^2 under *cay* and cay_{sim}, and it shows that the R^2 under cay_{sim} is comparable to the R^2 under *cay*. In other words, the irrelevant risk factor cay_{sim} can also dramatically improve the R^2. In this chapter, I explain why the result in Table 2.1 could happen, by deriving the distribution of the R^2 under irrelevant factors; in addition, I use the rank test of Kleibergen and Paap (2006) as a tool to investigate whether irrelevant risk factors commonly exist in empirical studies of the (C)CAPM.

Although the formal analysis of the R^2 is rare, the inadequacy of the FM two-pass procedure has been examined in several aspects. Kan and Zhang (1999) study a single irrelevant factor model, and provide theoretical results and simulation evidence to show the existence of bias in the second stage statistical tests and overflation in the cross-sectional R^2. Employing techniques from the weak identification literature, Kleibergen (2009) shows that the statistical inference based on the FM two-pass estimator is misleading when β is equal or close to zero. Lewellen *et al.* (2010) find that the R^2 of the FM two-pass procedure is not informative and offer suggestions to improve its performance, but their main point is not on the magnitude of β. This article extends this discussion by deriving the distribution of the R^2 of a linear multifactor pricing model, and the findings cast doubt on the success of the recent conditional versions of the (C)CAPM: an increase in the R^2 could be a byproduct of introducing irrelevant risk factors, hence are not strong evidence to support the conditional (C)CAPM.

I evaluate the performances of some recent versions of the (C)CAPM, including the following: Lettau and Ludvigson (2001), Lustig and Van Nieuwerburgh (2005), Yogo (2006), Santos and Veronesi (2006), Li *et al.* (2006). Previous studies indicate that risk factors proposed in these papers are successful in explaining the cross-sectional variation in asset returns. However, I can not rule out these factors are in fact irrelevant.

The rest of the chapter is organized as follows: the linear factor model is set up in Section 2; a rank test is suggested in Section 3; the distribution of the cross-sectional R^2 based on the FM two-pass procedure is derived in Section 4; in section 5, the performances of several versions of the (C)CAPM are evaluated; Section 6 concludes. The following notations are used throughout this chapter: " $\xrightarrow{d}$" indicates convergence in distribution, and " $\xrightarrow{p}$" indicates convergence in probability. For a $T \times n$ matrix $A = (a_1, ..., a_n)$, $P_A = A(A'A)^{-1}A'$, $M_A = I_T - P_A$, I_T is the $T \times T$ identity matrix, ι_T is the $T \times 1$ vector of ones, $vec(A) = (a_1', ..., a_n')'$, $vecinv_n((a_1', ..., a_n')') = A$, and $\otimes$ is the Kronecker operator.

2.2 Model and Fama-MacBeth

2.2.1 Linear Factor Model

The linear factor model is constructed by three equations:

$$E(R_t) = \iota_n \lambda_1 + \beta \lambda_F \tag{2.1}$$

$$cov(R_t, F_t) = \beta var(F_t) \tag{2.2}$$

$$E(F_t) = \mu_F \tag{2.3}$$

where excess asset returns R_t, market risk factors F_t, are n by 1, k by 1 vectors respectively.

Equation (2.1) reflects Sharpe (1964) and Lintner (1965)'s view: in equilibrium average returns should be priced by the risk measure β, plus the risk free return λ_1. Equation (2.2) defines β as the n by k correlation matrix of asset returns and risk factors. See Cochrane (2001) for a book-length discussion.

The unconditional version of CAPM considers the market return as the single factor, hence $k = 1$; in the unconditional CCAPM, the single factor is non-durable consumption growth. Fama and French (1992)(1993) show that a single factor model could only explain a small fraction of the total cross-sectional variation. The failings of the unconditional (C)CAPM have induced researchers to introduce new risk factors. For instance, a conditional version of the (C)CAPM adds at least a conditioning variable to F_t, hence $k \geq 2$. There exists a sizeable empirical literature on the (C)CAPM, suggesting various factors can successfully explain the cross-sectional variation of asset returns, and these factors include: the consumption-to-wealth ratio in Lettau and Ludvigson (2001), the labor income-to-consumption ratio in Santos and Veronesi (2006), the housing-collateral ratio in Lustig and Van Nieuwerburgh (2005), the investment growth rate in Li *et al.* (2006), the growth in durable consumption in Yogo (2006), the consumption risk in Parker and Julliard (2005), the housing expenditure in Piazzesi *et al.* (2007), etc.

Equations (2.1)(2.2)(2.3) imply the following model:

$$\begin{aligned} R_t &= \iota_n \lambda_1 + \beta \lambda_F + \beta v_t + u_t \\ &= \iota_n \lambda_1 + \beta(\bar{F}_t + \lambda_F) + \epsilon_t \\ F_t &= \mu_F + v_t \end{aligned}$$

where u_t, v_t are $n\times 1$, $k\times 1$ vectors of errors, $\bar{F}_t = F_t - \frac{1}{T}\sum_{t=1}^{T} F_t$, $\epsilon_t = u_t + \beta\frac{1}{T}\sum_{t=1}^{T} v_t$, and u_t, v_t are uncorrelated because of Equation (2.2).

The data observed by empirical researchers are $R_t, F_t, t = 1, ..., T$. To make expressions neat and facilitate the econometric analysis of the R^2, define matrices R, F, and a vector $\bar{R}$:

$$\begin{aligned} R_{n\times T} &= (R_1, R_2, ..., R_T) \\ F_{k\times T} &= (F_1, F_2, ..., F_T) \\ \bar{R}_{n\times 1} &= \frac{1}{T}\sum_{t=1}^{T} R_t \end{aligned}$$

2.2.2 Fama-MacBeth

The commonly used FM two-pass procedure in empirical studies of the (C)CAPM is as follows: (i) at the first stage, estimate β in a time series regression, i.e. regressing R_t on F_t with intercepts; (ii) estimate the risk premium λ_F using $\hat{\beta}$ in a second stage cross-sectional regression, i.e. regressing $\bar{R}$ on $\hat{\beta}$ with intercepts. Shanken (1992) provides a detailed econometric analysis of this approach. The expressions of $\hat{\beta}$, $\hat{\lambda}_F$, and the R^2 of the FM two-pass procedure are given below, where M_{ι_T}, M_{ι_n} are two projection matrices of constants (see Appendix A):

$$\hat{\beta} = RM_{\iota_T}F'(FM_{\iota_T}F')^{-1} \tag{2.4}$$

$$\hat{\lambda}_F = (\hat{\beta}'M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}'M_{\iota_n}\bar{R} \tag{2.5}$$

$$R^2 = \frac{\bar{R}'M_{\iota_n}\hat{\beta}(\hat{\beta}'M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}'M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}} \tag{2.6}$$

Given the setup of the model, a large value of the cross-sectional R^2 is expected

in empirical studies of the (C)CAPM. This is because the R^2 converges to 1 in large samples (see Appendix B), when the risk factors chosen by empirical researchers coincide with the factors in the above model. Consequently, a small value of the R^2 suggests that the chosen risk factors do not coincide with the factors in the model, hence are not useful in explaining the variation of asset returns.

2.3 A Rank Test

Note that the validity of the FM two-pass procedure relies on the full rank of the matrix β: in the second stage regression, the $n \times k$ matrix β is the matrix of regressors. If the rank of β is less than k, the rank condition of the ordinary least squares estimator is violated, and the validity of the FM two-pass procedure demonstrated by Shanken (1992) collapses.

When only a single factor is considered, $k = 1$, then the rank condition fails if this single factor is irrelevant, which is unlikely to happen. However, as the empirical literature on the (C)CAPM starts to propose various multifactor models, $k \geq 2$, there is a larger chance that β does not have full rank. For example, if one of the k factors is irrelevant, then a column of β is zero, hence its rank is reduced; more generally, if one column of β can be written as a linear combination of the other columns, then the rank condition fails to hold.

There exist several rank tests to examine the rank condition, given a consistent estimator $\hat{\beta}$ of the matrix β is available by the first-pass of the FM procedure: see Anderson (1951), Cragg and Donald (1996), etc. The rank test of Kleibergen and Paap (2006) is used in this chapter as this novel test overcomes some

deficiencies of other tests: it is robust to heteroscedasticity, while homoscedasticity is assumed in Anderson (1951); in addition, it is easier for implementation, while the rank test of Cragg and Donald (1996) involves numerical optimization.

If the rank test of Kleibergen and Paap (2006) suggests that β does not have full rank, it indicates the violation of the rank condition, and further casts doubt on the validity of the risk premium estimator and the cross-sectional R^2 from the second-pass of the FM procedure. Specifically, if an irrelevant factor is introduced by empirical researchers, then the rank of β in the empirical model is not full, hence the FM two-pass procedure is no longer valid; the rank test of Kleibergen and Paap (2006) helps determine whether this is the case.

Despite of its importance, the necessity of a rank test has not been recognized in empirical studies of the (C)CAPM, to the best of my knowledge. I now use the rank test of Kleibergen and Paap (2006) to show it is helpful for excluding irrelevant risk factors.

Take a specification of the conditional CAPM from Lettau and Ludvigson (2001) for example: the log consumption-wealth ratio *cay*, the value weighted return R_{vw}, and the interaction term $cay \cdot R_{vw}$ are the three factors. The estimate of β, the associated t statistic and p value using 25 size and book-to-market sorted portfolios are presented in Table 2.2. A feature observed from Table 2.2 that has motivated this chapter is that, R_{vw} is significantly related to all of the 25 portfolio returns, while *cay* is insignificant at 5% except for 2 portfolios only. The small magnitude of the estimate of β for *cay* induces the suspicion that the column of β corresponding to *cay* is zero, hence β does not have full rank. The result of the Kleibergen and Paap (2006) rank test is in line with Table 2.2: it tests the null that the rank of the β matrix (its dimension is 25 by 3 in this example) is 2, and reports a p value around

0.77. The large p value implies the failure of rejecting the null at 5% significance level, indicating that the rank condition is violated.

Table 2.3 presents the rank test results for three specifications of the (C)CAPM suggested in Lettau and Ludvigson (2001), together with the unconditional (C)CAPM, and the Fama-French three factor model. I can not reject that the three versions of the conditional (C)CAPM violate the rank condition, as their p values are large. In contrast, p values for the unconditional (C)CAPM and the Fama-French three factor model are all close to zero, indicating that the rank condition is satisfied for these models.

2.4 Distribution of R^2

As discussed above, if at least one risk factor is irrelevant, the rank condition is not satisfied. In this situation, the FM estimator of risk premium is inconsistent, as shown in Kleibergen (2009); however, it is not clear how the commonly used cross-sectional R^2 behaves. The presumption may be that irrelevant risk factors could not dramatically increase the R^2 at the second stage. This section shows this presumption is incorrect: the R^2 converges to a random variable if β does not have full rank, hence it could be large, even when factors are irrelevant.

Kan and Zhang (1999) have analytical results for the distribution of the R^2 under $k = 1$, which is generalized here to consider $k \geq 1$, because multifactor models are favored in recent empirical studies of the (C)CAPM. I am interested in answering the following question: if an irrelevant risk factor is used together with some relevant factors, how may the R^2 be affected?

To visualize the asymptotic distribution of R^2, Monte Carlo simulations are conducted to draw the densities of the risk premium estimator $\hat{\lambda}_F$ and the R^2. In the underlying data generating process (D.G.P.), $n = 25$ as the 25 Fama-French size and book-to-market sorted portfolios are the focus of many empirical studies. The risk premium λ_F is fixed to 1 if it exists, and the variance of errors are obtained from Lettau and Ludvigson (2001). 10000 replications are conducted. For simplicity, a three factor model is used in D.G.P. of the Monte Carlo study, unless otherwise stated.

2.4.1 Assumption

The same assumption as in Kleibergen (2009) is made. Assumption 1 is a statement of a central limit theorem, which implies that $\bar{R}$ and $\hat{\beta}$ follow two independent normal distributions as stated in Lemma 1. Proof of this lemma is contained in the appendix of Kleibergen (2009).

Assumption 1: $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\begin{pmatrix} 1 \\ F_t \end{pmatrix} \otimes (R_t - \iota_n \lambda_1 - \beta(\bar{F}_t + \lambda_F)) \right) \xrightarrow{d} \begin{pmatrix} \varphi_R \\ \varphi_\beta \end{pmatrix}$

Lemma 1: $\sqrt{T} \begin{pmatrix} \bar{R} - \iota_n \lambda_1 - \beta \lambda_F \\ vec(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \end{pmatrix}$

where $\psi_R(n \times 1), \psi_\beta(nk \times 1)$ are two independent normally distributed random variables with mean 0, and covariance matrices $\Omega, V_{FF}^{-1} \otimes \Omega$. $\Omega = var(\epsilon_t)$, $V_{FF} = var(F_t)$, $(\varphi_R', \varphi_\beta')' \sim N(0, V)$, $V = Q \otimes \Omega$, $Q = E\left[\begin{pmatrix} 1 \\ F_t \end{pmatrix} \begin{pmatrix} 1 \\ F_t \end{pmatrix}' \right]$.

Lemma 1 describes the distributions of $\bar{R}$ and $\hat{\beta}$, the main components of the R^2 in Equation (2.6). The asymptotic distribution of the R^2 is derived in terms of the distributions of $\bar{R}$ and $\hat{\beta}$.

2.4.2 Four Cases

I consider four cases, depending on the values of β.

Case 1: If $(\iota_n \vdots \beta)'(\iota_n \vdots \beta)$ has full rank, and risk factors used in the empirical study coincide with those in the model, then $R^2 \xrightarrow{p} 1$.

In this ideal setting, as the sample size T increases to ∞, $\hat{\lambda}_F$ converges to λ_F, and the R^2 goes to 1. See Figure 2.1$(a)(b)$. Figure 2.1$(a)(b)$ present the densities of $\hat{\lambda}_F$ and R^2 as the sample size T increases. Figure 2.1(a) shows that $\hat{\lambda}_F$ shrinks to the true value $\lambda_F = 1$, and Figure 2.1(b) shows that the cross-sectional R^2 converges to 1.

Case 1 illustrates the foundation of reporting the R^2 in practice: if the same risk factors F_t as in the underlying model are chosen, then the sample R^2 tends to be large. As a result, a large value of the cross-sectional R^2 is a positive indicator, since it suggests that the correct risk factors have been discovered. The R^2 thus serves as a criterion for model comparison in the empirical (C)CAPM literature: risk factors that yield larger R^2 are generally favored.

Not surprisingly, the way of using the R^2 to compare models has its pitfalls, many of which are documented in Lewellen *et al.* (2010). In this chapter, I focus on the problem induced by the magnitude of β: the second stage regression fails if β is

not large enough, and the R^2 consequently becomes a misleading measure. Case 2 describes an extreme case: useless factors can yield large values of the R^2.

Case 2: If $\beta = \mathbf{0}$ and $E(\bar{R}) = c$, then $R^2 \xrightarrow{d} \frac{c' P_{M_{\iota_n}\Psi_\beta} c}{c' M_{\iota_n} c}$. Specially, if $c = \iota_n \lambda_1$, then $R^2 \xrightarrow{d} \frac{\psi_R' P_{M_{\iota_n}\Psi_\beta} \psi_R}{\psi_R' M_{\iota_n} \psi_R}$, where $\Psi_\beta = vecinv_k(\psi_\beta)$, $P_{M_{\iota_n}\Psi_\beta} = M_{\iota_n}\Psi_\beta(\Psi_\beta' M_{\iota_n}\Psi_\beta)^{-1}\Psi_\beta' M_{\iota_n}$.

This is the extreme setting that all k factors included in the model are irrelevant, in the sense that these factors are uncorrelated with asset returns. Although it is tempting to think that in this setting the R^2 is zero as all factors have no power of pricing average asset returns, in fact the sample R^2 can be close to 1 even in large samples. See Figure 2.1$(c)(d)$. Figure 2.1(c) shows that $\hat{\lambda}_F$ does not converge to a point when the sample size increases, instead it stays randomly centered around 0. Similarly, Figure 2.1(d) shows that the R^2 does not converge to 0 or 1, but stays as a random variable, as Case 2 states.

The behaviors of the R^2 in Figure 2.1(b) and 2.1(d) are different: in Figure 2.1(d), the R^2 stays as a random variable, instead of converging to a fixed point. The randomness of the R^2 in the limiting distribution makes it a misleading criterion for model comparison, i.e. irrelevant factors have a positive probability of yielding large R^2 even in large samples, hence a model made of irrelevant factors always has a chance of being favored by empirical researchers, if the R^2 is the sole criterion of model comparison.

In practice, it is unlikely that all of the factors used by empirical researchers are unrelated with asset returns, hence the settings in Case 2 that β is zero for all k factors are very restrictive. What is possible, however, is that β is sizeable for most factors, but close to zero for one factor (or more). Case 3 corresponds to the setting

that a factor uncorrelated with asset returns is included in the model, together with several other relevant factors. The question under consideration is, if an irrelevant factor is introduced into the model as the conditioning variable, what will happen to the cross-sectional R^2 in the conditional (C)CAPM?

Case 3: Suppose $k_1 + 1$ factors with $\tilde{\beta} = (\beta_{k_1} \vdots \mathbf{0})$ are chosen in the empirical model, where β_{k_1} is $n \times k_1$, $\mathbf{0}$ is $n \times 1$, $k_1 < k$, $\sqrt{T}(\hat{\tilde{\beta}} - (\beta_{k_1} \vdots \mathbf{0})) \xrightarrow{d} \tilde{\psi}_\beta$, $\Psi_\beta = vecinv_{k_1+1}(\tilde{\psi}_\beta) = (\Psi_{k_1} \vdots \Psi_1)$, and $E(\bar{R}) = c$, then:

$$R^2 \xrightarrow{d} \frac{c' M_{\iota_n} (V_1 + V_2 + V_3 + V_4) M_{\iota_n} c}{c' M_{\iota_n} c}$$

where random variables V_1, V_2, V_3, V_4 are defined as follows:

$$\begin{aligned}
V_0 &\equiv (\beta'_{k_1} M_{\iota_n} \beta_{k_1} - \beta'_{k_1} M_{\iota_n} \Psi_1 (\Psi'_1 M_{\iota_n} \Psi_1)^{-1} \Psi'_1 M_{\iota_n} \beta_{k_1})^{-1} \\
V_1 &\equiv \beta_{k_1} V_0 \beta'_{k_1} \\
V_2 &\equiv -\beta_{k_1} V_0 \beta'_{k_1} M_{\iota_n} \Psi_1 (\Psi'_1 M_{\iota_n} \Psi_1)^{-1} \Psi'_1 \\
V_3 &\equiv -\Psi_1 (\Psi'_1 M_{\iota_n} \Psi_1)^{-1} \Psi'_1 M_{\iota_n} \beta_{k_1} V_0 \beta'_{k_1} \\
V_4 &\equiv \Psi_1 ((\Psi'_1 M_{\iota_n} \Psi_1)^{-1} + (\Psi'_1 M_{\iota_n} \Psi_1)^{-1} \Psi'_1 M_{\iota_n} \beta_{k_1} V_0 \beta'_{k_1} M_{\iota_n} \Psi_1 (\Psi'_1 M_{\iota_n} \Psi_1)^{-1}) \Psi'_1
\end{aligned}$$

This setting corresponds to the scenario that an irrelevant factor is added to an empirical model with other k_1 relevant factors. Case 3 states that adding an irrelevant factor causes the limiting distribution of the R^2 to be spurious. As shown by simulation, the value of the R^2 could be increased even if the added factor is irrelevant. See Figure 2.1$(e)(f)$. The Fama and French (1993) three factor model is used in the D.G.P. of the simulation for Figure 2.1$(e)(f)$. A single factor from the D.G.P., and an extra useless factor are chosen to compute the R^2 in Figure 2.1(f).

The useless factor is constructed by a normally distributed variate independent of other variables. Figure 2.1(e) plots the density of $\hat{\lambda}_F$ corresponding to the useless factor, and Figure 2.1(f) shows that the R^2 converges to a random variable after a useless factor is added.

The restriction in Case 3 can be loosened to allow the added factor to be weakly related to assets, i.e. β corresponding to this added factor is close, but not strictly equal to zero. This is considered in Case 4. The result is similar: the R^2 after including a nearly irrelevant factor is still random in its limiting distribution.

Case 4: Suppose $k_1 + 1$ factors with $\tilde{\beta} = (\beta_{k_1} \vdots \frac{b}{\sqrt{T}})$ are chosen in the empirical model, where β_{k_1} is $n \times k_1$, b is $n \times 1$, $k_1 < k$, $\sqrt{T}(\hat{\tilde{\beta}} - (\beta_{k_1} \vdots \frac{b}{\sqrt{T}})) \xrightarrow{d} \tilde{\psi}_\beta$, $\Psi_\beta = vecinv_{k_1+1}(\tilde{\psi}_\beta) = (\Psi_{k_1} \vdots \Psi_1)$, and $E(\bar{R}) = c$, then:

$$R^2 \xrightarrow{d} \frac{c' M_{\iota_n} (\tilde{V}_1 + \tilde{V}_2 + \tilde{V}_3 + \tilde{V}_4) M_{\iota_n} c}{c' M_{\iota_n} c}$$

where random variables $\tilde{V}_1$, $\tilde{V}_2$, $\tilde{V}_3$, $\tilde{V}_4$ are defined as follows:

$$\begin{aligned}
\tilde{\Psi}_1 &\equiv b + \Psi_1 \\
\tilde{V}_0 &\equiv (\beta_{k_1}' M_{\iota_n} \beta_{k_1} - \beta_{k_1}' M_{\iota_n} \tilde{\Psi}_1 (\tilde{\Psi}_1' M_{\iota_n} \tilde{\Psi}_1)^{-1} \tilde{\Psi}_1' M_{\iota_n} \beta_{k_1})^{-1} \\
\tilde{V}_1 &\equiv \beta_{k_1} \tilde{V}_0 \beta_{k_1}' \\
\tilde{V}_2 &\equiv -\beta_{k_1} \tilde{V}_0 \beta_{k_1}' M_{\iota_n} \tilde{\Psi}_1 (\tilde{\Psi}_1' M_{\iota_n} \tilde{\Psi}_1)^{-1} \tilde{\Psi}_1' \\
\tilde{V}_3 &\equiv -\tilde{\Psi}_1 (\Psi_1' M_{\iota_n} \tilde{\Psi}_1)^{-1} \tilde{\Psi}_1' M_{\iota_n} \beta_{k_1} \tilde{V}_0 \beta_{k_1}' \\
\tilde{V}_4 &\equiv \tilde{\Psi}_1 ((\Psi_1' M_{\iota_n} \tilde{\Psi}_1)^{-1} + (\tilde{\Psi}_1' M_{\iota_n} \tilde{\Psi}_1)^{-1} \tilde{\Psi}_1' M_{\iota_n} \beta_{k_1} \tilde{V}_0 \beta_{k_1}' M_{\iota_n} \tilde{\Psi}_1 (\tilde{\Psi}_1' M_{\iota_n} \tilde{\Psi}_1)^{-1}) \tilde{\Psi}_1'
\end{aligned}$$

Case 3 and 4 help explain the success of the conditional versions of the (C)CAPM. If the conditioning variable has β close to zero, then adding a conditioning variable

to the model is similar to adding an irrelevant factor, which causes the R^2 to fail to converge, hence it is not surprising to get large R^2's in empirical studies, even though risk factors are irrelevant. The motivating example presented in Table 2.1 in the introduction section makes the same point: irrelevant factors can appear useful in explaining the cross-sectional variation of asset returns.

To summarize, the analytical results in this section formalize the view that the R^2 is a misleading measure for the empirical studies of the (C)CAPM, if the rank condition on β is not satisfied. With irrelevant or nearly irrelevant factors, it is not surprising to find a sizeable value of the R^2, because the limiting distribution of the R^2 is spurious.

2.5 Examples of (C)CAPM

In the early part of this chapter, I show that for the specifications of the (C)CAPM in Lettau and Ludvigson (2001), the possibility of violating the rank condition can not be ruled out. Does this concern commonly exist in the empirical (C)CAPM literature, or is it just unique in Lettau and Ludvigson (2001)? To answer this question, I evaluate several versions of the (C)CAPM in this section by checking whether the rank condition is satisfied. As discussed above, a large p value of the Kleibergen and Paap (2006) rank test suggests the violation of the rank condition.

The 25 Fama-French size and book-to-market sorted portfolios are used as assets. Quarterly returns are compounded by monthly returns obtained from Kenneth French's web site, together with the Fama-French three factors, Rm-Rf, SMB, and HML. Details of constructing the portfolios and benchmark factors are available on

the web site.

I use the following factors: the nondurable consumption growth $\triangle c_{Ndur}$ and the durable consumption growth $\triangle c_{Dur}$ in Yogo (2006), the housing-collateral ratio $myfa$ in Lustig and Van Nieuwerburgh (2005), the labor income-to-consumption ratio s^w in Santos and Veronesi (2006), the investment growth rate in the household sector HHOLDS, the nonfinancial corporate business NFINCO, and the financial cooperations FINAN in Li *et al.* (2006). The data of these factors are either directly offered by the authors, or constructed following the descriptions in their paper. 1952Q1-2001Q4 is the time period within which all of the factors have data available.

One specification of the (C)CAPM from each paper listed above is used as an example in Table 4. The p value of the suggested rank test by Kleibergen and Paap (2006) is reported, together with estimation results based on the ordinary least squares (OLS), generalized least squares (GLS) (see Lewellen *et al.* (2010) for the GLS method).

The OLS R^2 in Table 4 is encouraging: all of the models explain at least 41% of the cross-sectional variation in average portfolio returns, when the FM two-pass procedure is applied. However, the results of the rank test are discouraging: except the benchmark Fama-French three factor model, all the other models have large p values.

The evaluation based on the GLS R^2 supports the outcome of the rank test: only the Fama-French three factor model has a large GLS R^2, while all the other models have the GLS R^2 under 20%. Compared to the OLS R^2, the GLS R^2 is much smaller for most models, except for the Fama-French model. These results hence support

the view in Lewellen *et al.* (2010): the GLS R^2 appears to be a better criterion than the OLS R^2.[1]

Overall, the violation of the rank condition can not be ruled out. Since introducing an irrelevant or nearly irrelevant factor into the linear multifactor model would make the R^2 spurious, as demonstrated in the last section, the large values of the OLS R^2 based on the FM two-pass procedure are not reliable.

2.6 Conclusion

This chapter cautions that if irrelevant risk factors are included in the empirical studies of the (C)CAPM, they may appear useful in explaining the cross-sectional variation of asset returns. With one or more irrelevant risk factors in a linear multifactor model, the full rank condition of the β matrix no longer holds, which further induces the spurious limiting behavior of the cross-sectional R^2. Consequently, the value of the sample R^2 can be large, even though factors are irrelevant.

From the perspective of the empirical (C)CAPM literature, this chapter highlights the necessity of applying a rank test on the β matrix, which helps make empirical results more reliable. The empirical findings are easy to summarize: I can not rule out the possibility that some recent versions of the (C)CAPM violate the rank condition, while the only model that remains trustworthy is the Fama and French (1993) three factor model.

[1]Lewellen, Nagel, and Shanken (2008) also use these factors as examples, and they use a different time period, 1963-2004, with 30 extra industry portfolios to expand the test assets.

Appendix

A. The cross-sectional R^2 based on the FM two-pass procedure

Equations (2.1)(2.2)(2.3) imply the following model:

$$\begin{aligned} R_t &= \iota_n \lambda_1 + \beta \lambda_F + \beta v_t + u_t \\ F_t &= \mu_F + v_t \end{aligned}$$

where u_t, v_t are $n \times 1, k \times 1$ vectors of errors, and u_t, v_t are uncorrelated. The model implies the following:

$$\begin{aligned} R_t &= \iota_n \lambda_1 + \beta(\lambda_F + F_t - \mu_F) + u_t \\ &= (\iota_n \lambda_1 + \beta \lambda_F - \beta \mu_F) + \beta F_t + u_t \end{aligned}$$

Use $R = (R_1, R_2, ..., R_T), F = (F_1, F_2, ..., F_T), U = (u_1, u_2, ..., u_T)$:

$$\begin{aligned} R &= \iota_T' \otimes (\iota_n \lambda_1 + \beta \lambda_F - \beta \mu_F) + \beta F + U \\ M_{\iota_T} R' &= M_{\iota_T} F' \beta' + M_{\iota_T} U' \end{aligned}$$

Hence the OLS estimator $\hat{\beta}$ is:

$$\hat{\beta} = [(F M_{\iota_T} F')^{-1} F M_{\iota_T} R']' = R M_{\iota_T} F' (F M_{\iota_T} F')^{-1}$$

Similary, $\hat{\lambda}_F = (\hat{\beta}' M_{\iota_n} \hat{\beta})^{-1} \hat{\beta}' M_{\iota_n} \bar{R}$, where $\bar{R} = \frac{1}{T} \Sigma_{t=1}^{T} R_t$. By definition:

$$\begin{aligned} R^2 &= \frac{(M_{\iota_n}\hat{\beta}\hat{\lambda}_F)'(M_{\iota_n}\hat{\beta}\hat{\lambda}_F)}{(M_{\iota_n}\bar{R})'(M_{\iota_n}\bar{R})} \\ &= \frac{\hat{\lambda}_F'\hat{\beta}' M_{\iota_n}\hat{\beta}\hat{\lambda}_F}{\bar{R}' M_{\iota_n}\bar{R}} \\ &= \frac{\bar{R}' M_{\iota_n}\hat{\beta}(\hat{\beta}' M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}' M_{\iota_n}\hat{\beta}(\hat{\beta}' M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}' M_{\iota_n}\bar{R}}{\bar{R}' M_{\iota_n}\bar{R}} \\ &= \frac{\bar{R}' M_{\iota_n}\hat{\beta}(\hat{\beta}' M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}' M_{\iota_n}\bar{R}}{\bar{R}' M_{\iota_n}\bar{R}} \\ &= \frac{\bar{R}' P_{M_{\iota_n}\hat{\beta}}\bar{R}}{\bar{R}' M_{\iota_n}\bar{R}} \end{aligned}$$

where $P_{M_{\iota_n}\hat{\beta}} = M_{\iota_n}\hat{\beta}(\hat{\beta}' M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}' M_{\iota_n}$.

B. **Proof of Case 1**

Start from Equation (2.1):

$$\begin{aligned} E(R_t) &= \iota_n\lambda_1 + \beta\lambda_F \\ M_{\iota_n}E(R_t) &= M_{\iota_n}\iota_n\lambda_1 + M_{\iota_n}\beta\lambda_F \\ M_{\iota_n}E(R_t) &= M_{\iota_n}\beta\lambda_F \\ \frac{(M_{\iota_n}\beta\lambda_F)'(M_{\iota_n}\beta\lambda_F)}{(M_{\iota_n}E(R_t))'(M_{\iota_n}E(R_t))} &= 1 \end{aligned}$$

In the ideal setting, as $T \to \infty$, $\hat{\beta} \xrightarrow{p} \beta$, $\hat{\lambda}_F \xrightarrow{p} \lambda_F$ and $\bar{R} \xrightarrow{p} E(R_t)$, by Slutsky's theorem:

$$\begin{aligned} R^2 &= \frac{(M_{\iota_n}\hat{\beta}\hat{\lambda}_F)'(M_{\iota_n}\hat{\beta}\hat{\lambda}_F)}{(M_{\iota_n}\bar{R})' M_{\iota_n}\bar{R}} \\ &\xrightarrow{p} \frac{(M_{\iota_n}\beta\lambda_F)'(M_{\iota_n}\beta\lambda_F)}{(M_{\iota_n}E(R_t))'(M_{\iota_n}E(R_t))} \\ &\xrightarrow{p} 1 \end{aligned}$$

C. **Proof of Case 2**

If $\beta = 0$ and $E(\bar{R}) = c$, Lemma 1 reduces to:

$$\sqrt{T}\begin{pmatrix} \bar{R}-c \\ vec(\hat{\beta}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \end{pmatrix}$$

Hence, the following are true:

$$\begin{aligned} M_{\iota_n}\bar{R} &\xrightarrow{p} M_{\iota_n}c \\ \sqrt{T}\hat{\beta} &\xrightarrow{d} vecinv_k(\psi_\beta) = \Psi_\beta \end{aligned}$$

Apply the continuous mapping theorem:

$$\begin{aligned} R^2 &= \frac{\bar{R}'M_{\iota_n}\hat{\beta}(\hat{\beta}'M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}'M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}} \\ &\xrightarrow{d} \frac{c'M_{\iota_n}\Psi_\beta(\Psi_\beta'M_{\iota_n}\Psi_\beta)^{-1}\Psi_\beta'M_{\iota_n}c}{c'M_{\iota_n}c} \\ &\xrightarrow{d} \frac{c'P_{M_{\iota_n}\Psi_\beta}c}{c'M_{\iota_n}c} \end{aligned}$$

As a special case, if $c = \iota_n\lambda_1$, then:

$$\begin{aligned} \sqrt{T}M_{\iota_n}\bar{R} &= \sqrt{T}M_{\iota_n}(\bar{R}-\iota_n\lambda_1) \\ &\xrightarrow{d} M_{\iota_n}\psi_R \end{aligned}$$

Apply the continuous mapping theorem:

$$\begin{aligned} R^2 &= \frac{\bar{R}'M_{\iota_n}\hat{\beta}(\hat{\beta}'M_{\iota_n}\hat{\beta})^{-1}\hat{\beta}'M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}} \\ &\xrightarrow{d} \frac{\psi_R'M_{\iota_n}\Psi_\beta(\Psi_\beta'M_{\iota_n}\Psi_\beta)^{-1}\Psi_\beta'M_{\iota_n}\psi_R}{\psi_R'M_{\iota_n}\psi_R} \\ &\xrightarrow{d} \frac{\psi_R'P_{M_{\iota_n}\Psi_\beta}\psi_R}{\psi_R'M_{\iota_n}\psi_R} \end{aligned}$$

D. **Proof of Case 3**

If $\tilde{\beta} = (\beta_{k_1} : \mathbf{0})$ and $E(\bar{R}) = c$, Lemma 1 reduces to:

$$\sqrt{T}\begin{pmatrix} \bar{R}-c \\ vec(\hat{\tilde{\beta}}-(\beta_{k_1}:\mathbf{0})) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \tilde{\psi}_\beta \end{pmatrix}$$

Hence, the following are true:

$$\begin{aligned} \bar{R} &\xrightarrow{p} c \\ \sqrt{T}(\hat{\tilde{\beta}}-(\beta_{k_1}:\mathbf{0})) &\xrightarrow{d} \Psi_\beta = (\Psi_{k_1}:\Psi_1) \\ \hat{\tilde{\beta}} &= (\beta_{k_1}+\Psi_{k_1}/\sqrt{T}+o(T^{-\frac{1}{2}}):\Psi_1/\sqrt{T}+o(T^{-\frac{1}{2}})) \end{aligned}$$

This gives the expression of R^2 as:

$$\begin{aligned} R^2 &= \frac{\bar{R}'M_{\iota_n}\hat{\tilde{\beta}}(\hat{\tilde{\beta}}'M_{\iota_n}\hat{\tilde{\beta}})^{-1}\hat{\tilde{\beta}}'M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}} \\ &= \frac{\bar{R}'M_{\iota_n}\hat{\tilde{\beta}}\begin{pmatrix} \beta_{k_1}'M_{\iota_n}\beta_{k_1}+O(T^{-\frac{1}{2}}) & \beta_{k_1}'M_{\iota_n}\Psi_1/\sqrt{T}+o(T^{-\frac{1}{2}}) \\ \Psi_1'M_{\iota_n}\beta_{k_1}/\sqrt{T}+o(T^{-\frac{1}{2}}) & \Psi_1'M_{\iota_n}\Psi_1/T+o(T^{-1}) \end{pmatrix}^{-1}\hat{\tilde{\beta}}'M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}} \end{aligned}$$

The formula of inverse of a block matrix states, with $S_D = A - BD^{-1}C$:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} S_D^{-1} & -S_D^{-1}BD^{-1} \\ -D^{-1}CS_D^{-1} & D^{-1}+D^{-1}CS_D^{-1}BD^{-1} \end{pmatrix}$$

Introduce new notations to make expressions neat:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^* & B^* \\ C^* & D^* \end{pmatrix}$$

Apply the formula and rearrange terms:

$$R^2 = \frac{\bar{R}'M_{\iota_n}(\beta_{k_1}A^*\beta_{k_1}'+\beta_{k_1}B^*\Psi_1'/\sqrt{T}+\Psi_1C^*\beta_{k_1}'/\sqrt{T}+\Psi_1D^*\Psi_1'/T+O(T^{-\frac{1}{2}}))M_{\iota_n}\bar{R}}{\bar{R}'M_{\iota_n}\bar{R}}$$

Derive asymptotic distributions of the four leading terms above:

$$\begin{aligned}
A^* &\xrightarrow{d} (\beta_{k_1}' M_{\iota_n} \beta_{k_1} - \beta_{k_1}' M_{\iota_n} \Psi_1 (\Psi_1' M_{\iota_n} \Psi_1)^{-1} \Psi_1' M_{\iota_n} \beta_{k_1})^{-1} \equiv V_0 \\
\beta_1 A^* \beta_{k_1}' &\xrightarrow{d} \beta_{k_1} V_0 \beta_{k_1}' \equiv V_1 \\
\beta_1 B^* \Psi_1' / \sqrt{T} &\xrightarrow{d} -\beta_{k_1} V_0 \beta_{k_1}' M_{\iota_n} \Psi_1 (\Psi_1' M_{\iota_n} \Psi_1)^{-1} \Psi_1' \equiv V_2 \\
\Psi_1 C^* \beta_1' / \sqrt{T} &\xrightarrow{d} -\Psi_1 (\Psi_1' M_{\iota_n} \Psi_1)^{-1} \Psi_1' M_{\iota_n} \beta_{k_1} V_0 \beta_{k_1}' \equiv V_3 \\
\Psi_1 D^* \Psi_1' / T &\xrightarrow{d} \Psi_1 (\Psi_1' M_{\iota_n} \Psi_1)^{-1} (1 + \Psi_1' M_{\iota_n} \beta_{k_1} V_0 \beta_{k_1}' M_{\iota_n} \Psi_1 (\Psi_1' M_{\iota_n} \Psi_1)^{-1}) \Psi_1' \equiv V_4
\end{aligned}$$

Apply the continuous mapping theorem:

$$R^2 \xrightarrow{d} \frac{c' M_{\iota_n} (V_1 + V_2 + V_3 + V_4) M_{\iota_n} c}{c' M_{\iota_n} c}$$

E. **Proof of Case 4**

If $\tilde{\beta} = (\beta_{k_1} : \frac{b}{\sqrt{T}})$ and $E(\bar{R}) = c$, Lemma 1 reduces to:

$$\sqrt{T} \begin{pmatrix} \bar{R} - c \\ vec(\hat{\tilde{\beta}} - (\beta_{k_1} : \frac{b}{\sqrt{T}})) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \tilde{\psi}_\beta \end{pmatrix}$$

Hence, the following are true:

$$\begin{aligned}
\bar{R} &\xrightarrow{p} c \\
\sqrt{T}(\hat{\tilde{\beta}} - (\beta_{k_1} : \frac{b}{\sqrt{T}})) &\xrightarrow{d} \Psi_\beta = (\Psi_{k_1} : \Psi_1) \\
\hat{\tilde{\beta}} &= (\beta_{k_1} + \Psi_{k_1}/\sqrt{T} + o(T^{-\frac{1}{2}}) : (b + \Psi_1)/\sqrt{T} + o(T^{-\frac{1}{2}}))
\end{aligned}$$

Define $\tilde{\Psi}_1 \equiv b + \Psi_1$, and follow the steps in the proof of Case 3.

Table 2.1: A motivating example, R^2 under cay and cay_{sim}

	cay	cay_{sim}
Specification (4)	0.31	0.44
(5)	0.31	0.29
(6)	0.77	0.72
(7)	0.75	0.69

Notes: This table presents the values of the cross-section R^2 of the Fama-MacBeth two-pass procedure using the versions of the (C)CAPM in Lettau and Ludvigson (2001). R^2's under cay are identical to those reported in Table 1 of Lettau and Ludvigson (2001), and the results are replicated using the same data as in Lettau and Ludvigson (2001). R^2's under cay_{sim} are the average of 10000 simulations: cay_{sim} is randomly drawn from a normal distribution, with the same mean and variance as cay. In each simulation, we compute the R^2 replacing cay with cay_{sim}, and take the average of 10000 simulations. Each row in this table corresponds to a specification of the (C)CAPM in Table 1 of Lettau and Ludvigson (2001). For instance, Specification (6) corresponds to a five-factor version of the (C)CAPM: the log consumption-wealth ratio cay, the value weighted return R_{vw}, the labor income growth $\triangle y$, and the interaction terms $cay \cdot R_{vw}$, $cay \cdot \triangle y$; in the simulation, we use cay_{sim}, R_{vw}, $\triangle y$, $cay_{sim} \cdot R_{vw}$, $cay_{sim} \cdot \triangle y$ to generate the R^2 under cay_{sim}. This table shows that the irrelevant cay_{sim} can increase the R^2, just like cay does.

Table 2.2: An example of the β matrix in Lettau and Ludvigson (2001)

	cay			R_{vw}			$cay \cdot R_{vw}$		
25 portfolios	β_{cay}	t	p	$\beta_{R_{vw}}$	t	p	$\beta_{cay \cdot R_{vw}}$	t	p
1	0.02	0.30	0.76	1.60	17.17	0	0.21	0.31	0.76
2	0.03	0.47	0.64	1.43	17.48	0	0.27	0.45	0.66
3	0.02	0.30	0.76	1.32	17.64	0	0.31	0.56	0.58
4	0.02	0.31	0.76	1.25	16.93	0	0.08	0.14	0.89
5	-0.01	-0.09	0.93	1.28	14.74	0	0.60	0.92	0.36
6	0.03	0.56	0.57	1.54	21.81	0	0.07	0.14	0.89
7	0.02	0.35	0.73	1.35	21.97	0	-0.20	-0.44	0.66
8	0.01	0.30	0.76	1.24	22.99	0	0.00	0.01	1.00
9	0.02	0.63	0.53	1.14	21.44	0	0.29	0.73	0.47
10	-0.04	-0.83	0.41	1.18	17.49	0	0.65	1.29	0.20
11	0.04	1.04	0.30	1.43	26.61	0	-0.32	-0.79	0.43
12	0.07	2.21	0.03	1.21	27.24	0	0.02	0.05	0.96
13	0.03	0.87	0.39	1.10	25.72	0	0.02	0.06	0.95
14	0.03	0.82	0.41	1.03	23.73	0	0.47	1.46	0.15
15	-0.02	-0.47	0.64	1.09	17.49	0	-0.01	-0.02	0.99
16	0.06	1.95	0.05	1.25	30.45	0	-0.27	-0.89	0.37
17	0.03	1.05	0.29	1.16	35.03	0	-0.33	-1.35	0.18
18	-0.01	-0.30	0.77	1.05	31.86	0	0.27	1.10	0.27
19	-0.01	-0.48	0.63	1.02	25.52	0	0.75	2.50	0.01
20	-0.04	-0.95	0.34	1.11	19.35	0	1.05	2.44	0.01
21	-0.05	-2.09	0.04	1.04	30.46	0	0.30	1.17	0.24
22	0.02	0.89	0.37	0.95	33.89	0	0.19	0.91	0.36
23	0.04	1.44	0.15	0.77	22.93	0	-0.25	-1.01	0.31
24	-0.01	-0.36	0.72	0.83	24.80	0	0.10	0.40	0.69
25	-0.04	-0.96	0.34	0.80	14.81	0	0.16	0.39	0.70

Notes: This table presents the estimate of the correlation matrix $\beta = (\beta_{cay}, \beta_{R_{vw}}, \beta_{cay \cdot R_{vw}})$, and its associated t statistic and p value, from the first stage time series regression of the Fama-MacBeth two-pass procedure. The three risks factors are the log consumption-wealth ratio cay, the value weighted return R_{vw}, and the interaction term $cay \cdot R_{vw}$ in Lettau and Ludvigson (2001). The quarterly returns of 25 Fama-French size and book-to-market sorted portfolios are from $1963Q3$ to $1998Q3$.

Table 2.3: p value of the rank test for 6 specifications in Lettau and Ludvigson (2001)

	cay	R_{vw}	$\triangle y$	SMB	HML	cay·R_{vw}	cay·$\triangle y$	$\triangle c$	cay·$\triangle c$	R^2	p value
1		-0.32								0.01	0.00
		(0.96)									
2								0.22		0.16	0.00
								(0.18)			
3		1.33		0.47	1.46					0.80	0.00
		(1.59)		(0.10)	(0.12)						
4	-0.52	-0.06				1.14				0.31	0.77
	(3.28)	(1.47)				(0.46)					
5	-0.44	-1.99	0.56			0.34	-0.17			0.77	0.55
	(0.45)	(1.58)	(0.44)			(0.32)	(0.12)				
6	-0.13							0.02	0.06	0.70	0.33
	(0.40)							(0.16)	(0.02)		

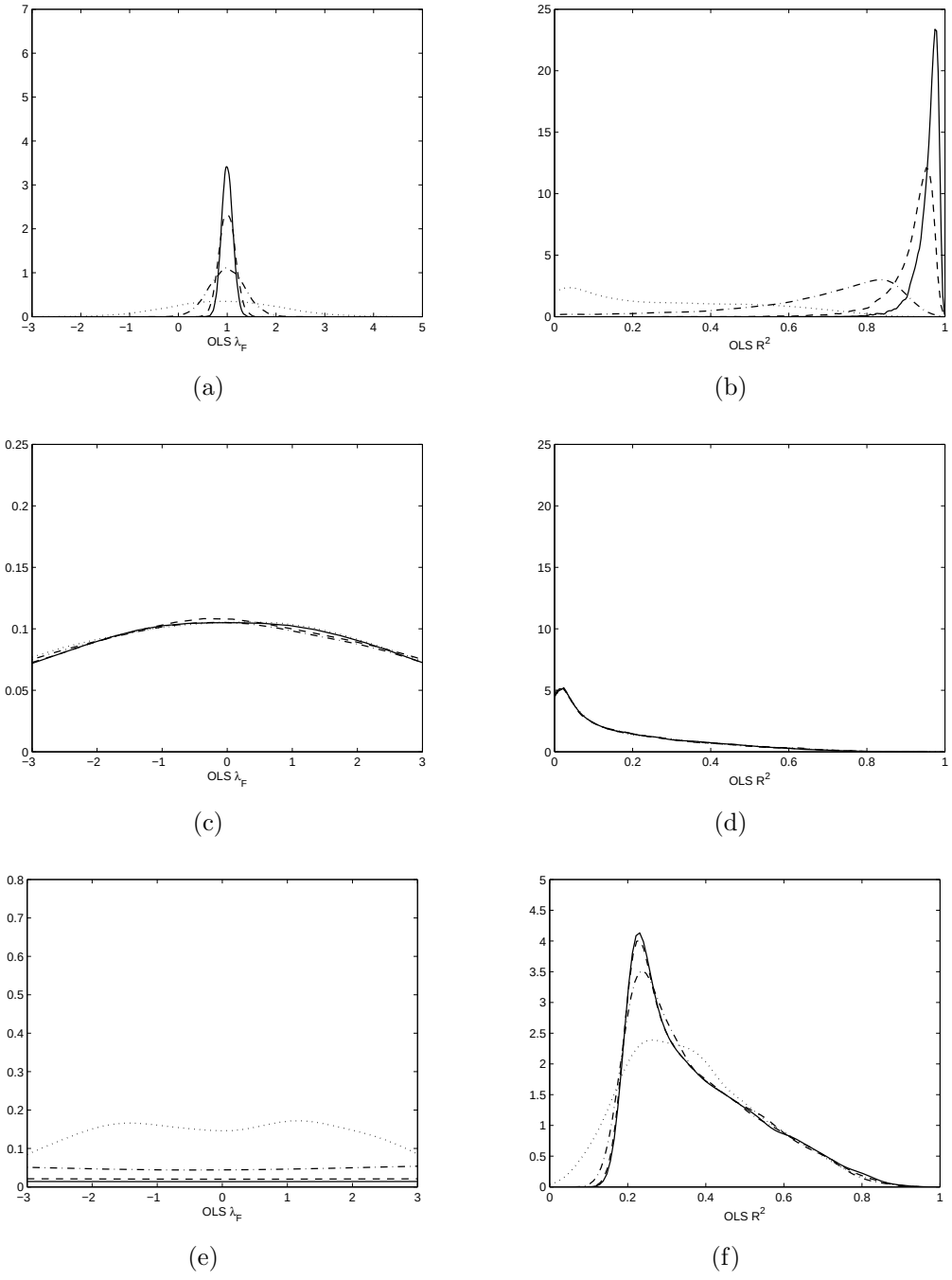
Notes: This table presents p values of the rank test in Kleibergen and Paap (2006), for 6 specifications of the (C)CAPM. A large p value indicates that the rank condition is likely violated. These 6 specifications are: the unconditional CAPM is Specification 1, the unconditional CCAPM is Specification 2, the Fama-French three factor model is Specification 3, and the other three specifications are from Lettau and Ludvigson (2001). We use the same data as Lettau and Ludvigson (2001), hence the OLS estimates of the risk premium and the R^2 are identical to Lettau and Ludvigson (2001). Standard errors with Shanken (1992) correction are in brackets. The quarterly returns of 25 Fama-French size and book-to-market sorted portfolios are from 1963Q3 to 1998Q3.

Table 2.4: p value of the rank test for 5 specifications of (C)CAPM

Model	Factors			R^2	p value
Fama-French (1993)	R_M	SMB	HML		
OLS	-1.06	0.46	1.34	0.73	0.00
	(1.30)	(0.11)	(0.15)		
GLS	-1.50	0.59	1.06	0.79	
	(0.98)	(0.06)	(0.11)		
Lustig-Nieuwerburgh (2005)	$myfa$	$\triangle c$	$myfa\triangle c$		
OLS	0.02	0.14	0.05	0.74	0.54
	(0.03)	(0.33)	(0.03)		
GLS	-0.01	-0.05	0.02	0.18	
	(0.01)	(0.12)	(0.01)		
Li-Vassalou-Xing (2006)	HHOLDS	NFINCO	FINAN		
OLS	0.03	0.01	0.09	0.58	0.57
	(0.02)	(0.01)	(0.04)		
GLS	0.01	-0.01	0.01	0.19	
	(0.01)	(0.01)	(0.02)		
Santos-Veronesi (2006)	R_M	s^w			
OLS	0.53	-2.38		0.41	0.96
	(1.73)	(1.14)			
GLS	-0.99	-0.15		0.04	
	(0.68)	(0.29)			
Yogo (2006)	R_M	$\triangle c_{Ndur}$	$\triangle c_{Dur}$		
OLS	0.06	0.68	-0.12	0.54	0.78
	(1.70)	(0.38)	(0.30)		
GLS	-1.03	-0.01	-0.03	0.03	
	(0.69)	(0.10)	(0.12)		

Notes: The table presents the OLS, GLS estimates of risk premium, the R^2, and p value from the rank test of Kleibergen and Paap (2006). 5 specifications of the (C)CAPM are considered, including the benchmark Fama-French three factor model. For each specification, the OLS, GLS estimates of the risk premium with the corresponding R^2 separated by rows are reported: OLS in the first row, GLS in the second row. Standard errors with Shanken (1992) correction are in brackets. The quarterly returns of 25 Fama-French size and book-to-market sorted portfolios used for this table are from 1952Q1 to 2001Q4.

Figure 2.1: Densities of $\hat{\lambda}_F$ and R^2



Notes: This figure presents the densities of $\hat{\lambda}_F$ and R^2 as the sample size T increases. $T = 1400$(dotted), 14000(dash-dotted), 70000(dashed), 140000(solid). Case 1 is in Row 1, where factors are correctly chosen; Case 2 is in Row 2, where all factors are irrelevant; Case 3 is in Row 3, where one irrelevant factor is chosen with one relevant factor.

CHAPTER **Three**

Does a Technology Shock Increase or Decrease Hours

3.1 Introduction

[1]Structural Vector Autoregressions (SVAR) have recently been employed to investigate the impact of technology shocks on production inputs, e.g. the hours worked. The empirical findings are conflicting, and there is currently a debate on whether a positive technology shock would increase or decrease hours: for example, Gali (1999), Shea (1999), Francis and Ramey (2005) find that a positive technology shock decreases hours at short horizons, while Christiano *et al.* (2003) report the increase in hours after a positive technology shock. The sign of the impact of technology shocks on hours is of interest, because real business cycle models have the typical implication that the hours worked will increase after a positive shock to technology, while the implication of models with sticky prices is often the opposite: see, e.g. Rebelo (2005).

The widely used identification strategy of the aforementioned empirical studies in SVAR is to impose the same restriction that only technology shocks have the long run effect on labor productivity, see e.g., Gali (1999). In addition, it is found that empirical outcome crucially depends on whether the time series of hours are specified in levels or in first differences: as stated in Pesavento and Rossi (2005), the decrease in hours is typically found in the first difference specification of hours, while the increase is often found in the level specification. Since the series of hours is highly persistent, and unit root tests are often not powerful enough to determine which specification should be chosen, Pesavento and Rossi (2005), Gospodinov (2010) etc. model hours as local-to-unity, an instrument to nest both the level and difference specifications. However, it has been shown that identification through the long-run restriction becomes weak when the variables that enter the model are highly

[1]This chapter is based on a joint project with Sophocles Mavroeidis.

persistent, and under the local-to-unity asymptotics, Gospodinov (2010) proves that the structural parameters as well as the impulse response functions (IRF) of interest can not be consistently estimated: this failure in estimation is interpreted as a weak identification problem in Gospodinov (2010), by using the framework of the linear instrumental variable (IV) regression in Staiger and Stock (1997).

From the weak identification perspective, although a consistent point estimator may not exist, it is often feasible to construct a confidence interval for the parameter of interest. In this chapter, we propose two tests to construct such confidence intervals, and we consider these intervals as *robust*, because no matter whether the true specification of hours is in levels or in first differences, these intervals would cover the structural parameter of interest with the probability at least as high as the nominal coverage rate. Given the robust intervals for structural parameters are derivable, and the IRF is a function of structural parameters, we consequently construct the bounds for the IRF to investigate whether a positive technology shock would increase or decrease the hours worked. The empirical findings are as follows: the impact of a positive technology shock on hours is likely positive at short horizons, but the 95% confidence interval of the contemporaneous effect also includes a negative region.

The rest of the chapter contains the following parts: the VAR($p+1$) model is described in Section 2; our robust approach of deriving the confidence intervals are presented in Section 3; Section 4 contains the empirical findings; Section 5 concludes.

3.2 Model

Following the convention, we use the following notations throughout this chapter: let l_t, h_t denote labor productivity and the hours worked respectively; structural shocks $\epsilon_t = (\epsilon_t^z, \epsilon_t^d)'$, where ϵ_t^z is the technology shock and ϵ_t^d is the non-technology shock; the object of interest is the IRF $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$, i.e. the response of hours to positive technology shocks, and we would like to explore whether its sign is positive or negative for small j's; in the special case that $j = 0$, $\frac{\partial h_t}{\partial \epsilon_t^z}$ is the contemporaneous effect of the technology shock on the hours worked.

3.2.1 VAR

We consider the same setup as in Gospodinov *et al.* (2009), which is an extension of the model in Gospodinov (2010). $\tilde{Y}_t = \begin{pmatrix} l_t \\ h_t \end{pmatrix}$, $t = 1, ..., T$, are assumed to be generated by a bivariate VAR$(p+1)$ model of (3.1) with assumptions $i - iv$:

$$\Psi(L)(I - \Phi L)\tilde{Y}_t \;=\; u_t \tag{3.1}$$

where $\Psi(L) = I - \sum_{i=1}^{p} \Psi_i L^i = \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix}, \Phi = \begin{pmatrix} 1 & \beta(\rho - 1) \\ 0 & \rho \end{pmatrix}$.

i. $u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$ is i.i.d. with covariance $\Sigma_u = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ and finite fourth moments;

ii. the largest roots of the system are contained in Φ with the conditions that l_t has a unit-root, while h_t follows a local-to-unity process, i.e. $\rho = 1 + \frac{c}{T}$,

and c is a fixed constant: this device nests both the first difference and level specifications of h_t, depending on whether $c = 0$ or $c < 0$;

iii. $|\Psi(z)| = 0$ has roots outside the unit circle and $\Psi(z)^{-1}$ is one-summable;

iv. there is an off diagonal element $\beta(\rho - 1)$ in Φ.

Gospodinov *et al.* (2009) and Gospodinov (2010) both emphasize that it is appropriate to use the model of (3.1) to study the impact of technology shocks on hours, and the off diagonal element $\beta(\rho - 1)$ in Φ is crucial: $\beta \neq 0$ allows the low frequency co-movement between l_t and h_t in the level specification of h_t, and this co-movement is assumed to be removed by the first difference filter, i.e. when $c = 0$, Φ reduces to an identity matrix. Gospodinov *et al.* (2009) offer arguments for why $\beta(\rho - 1)$ needs to be in place. For example, if technology shocks have lasting effects on the labor market, then the low frequency co-movement between l_t and h_t may be plausible. Following Gospodinov *et al.* (2009), the lower-left element in Φ is set to zero, to avoid that h_t is $I(2)$ when $c = 0$ or $I(1)$ when $c \neq 0$. If $\beta = 0$, (3.1) reduces to the model considered in Gospodinov (2010), hence (3.1) is an extension of the setup in Gospodinov (2010).

3.2.2 Long Run Restriction

Rewrite (3.1) as $A(L)Y_t = u_t$, with $A(L) = \Psi(L) \begin{pmatrix} 1 & \beta(1-\rho)L \\ 0 & 1-\rho L \end{pmatrix}$, $Y_t = \begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix}$.

$A(L)Y_t = u_t$ is thus equivalent to:

$$\begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} 1 & \beta(1-\rho)L \\ 0 & 1-\rho L \end{pmatrix} \begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix} = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

Premultiplying the equation above by $B_0 = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}$ yields the SVAR $B(L)Y_t = \epsilon_t$, with $B(L) = B_0 A(L)$, $\epsilon_t = B_0 u_t$:

$$\begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix} \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} 1 & \beta(1-\rho)L \\ 0 & 1-\rho L \end{pmatrix} \begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix} = \begin{pmatrix} \epsilon_t^z \\ \epsilon_t^d \end{pmatrix}$$

The identification restriction that ϵ_t^d has no long run effect on l_t implies two different expressions for the structural parameter b_{12}, depending on whether $c = 0$ or not:

i. if $c \neq 0$:
   Let $M_1(L) = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix} \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} 1 & \beta(1-\rho)L \\ 0 & 1-\rho L \end{pmatrix}$, then the long run restriction corresponds to that $M_1(1)$ is lower-triangular, i.e. the upper-right element of the 2 by 2 matrix $M_1(1)$ is zero, implying $b_{12} = \frac{\psi_{12}(1)+\beta\psi_{11}(1)}{\psi_{22}(1)+\beta\psi_{21}(1)}$;

ii. if $c = 0$:
   Let $M_2(L) = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix} \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix}$, then the long run restric-

tion corresponds to that the upper-right element in $M_2(1)$ is zero, implying $b_{12} = \frac{\psi_{12}(1)}{\psi_{22}(1)}$.

Gospodinov *et al.* (2009) use the above discontinuity in the solution of b_{12} to explain why the level specification and the first difference specification of h_t produce substantially different IRF's in empirical studies: the IRF's are functions of the structural parameters; using a different specification of h_t implies a different identification condition for b_{12}, as described above, hence it is not surprising that IRFs substantially differ when the specifications of h_t differ, if $\beta \neq 0$.

The condition $\beta \neq 0$ is thus crucial to help reconcile the conflicting empirical results: the discontinuity in the solution of b_{12} disappears once $\beta = 0$ is imposed, i.e. $b_{12} = \frac{\psi_{12}(1)}{\psi_{22}(1)}$, no matter whether $c = 0$ or $c \neq 0$. Gospodinov (2010) proposes to infer b_{12} by $\frac{\psi_{12}(1)}{\psi_{22}(1)}$, after imposing $\beta = 0$; without imposing $\beta = 0$, Gospodinov (2010)'s approach of deriving b_{12} through $\frac{\psi_{12}(1)}{\psi_{22}(1)}$ is no longer applicable in the current setup. In the later part of this chapter, we will propose methods to construct C.I.'s for b_{12}, without assuming $\beta = 0$.

3.2.3 Model Simplification

Define $\tilde{A}(L) = \Psi(L)(I - \Phi L)$, $A^*(L) = L^{-1}(I - \tilde{A}(L)) = A_0^* + A_1^* L + ... + A_p^* L^p$, $A_i^{**} = -\sum_{j=i+1}^{p} A_j^*$.

The Vector Error Correction (VEC) form of (3.1) is as follows:

$$\triangle \tilde{Y}_t = (A^*(1) - I)\tilde{Y}_{t-1} + \sum_{i=1}^{p} A_{i-1}^{**} \triangle \tilde{Y}_{t-i} + u_t$$

More explicitly:

$$\begin{cases} \triangle l_t = (\psi_{12}(1) + \beta\psi_{11}(1))\frac{c}{T}h_{t-1} + \sum_{i=1}^{p} a^{**}_{i-1,11}\triangle l_{t-i} + \sum_{i=1}^{p} a^{**}_{i-1,12}\triangle h_{t-i} + u_{1,t} \\ \triangle h_t = (\psi_{22}(1) + \beta\psi_{21}(1))\frac{c}{T}h_{t-1} + \sum_{i=1}^{p} a^{**}_{i-1,21}\triangle l_{t-i} + \sum_{i=1}^{p} a^{**}_{i-1,22}\triangle h_{t-i} + u_{2,t} \end{cases}$$

To simplify the VEC form above, project $\triangle l_t$, $\triangle h_t$ and h_{t-1} on the predetermined variables $\triangle l_{t-1}$, ..., $\triangle l_{t-p}$, $\triangle h_{t-1}$, ..., $\triangle h_{t-p}$, and save the residuals from projection as $\triangle\tilde{l}_t$, $\triangle\tilde{h}_t$ and $\tilde{h}_{t-1}$, i.e. regress $\triangle l_t$, $\triangle h_t$ and h_{t-1} on $\triangle l_{t-1}$, ..., $\triangle l_{t-p}$, $\triangle h_{t-1}$, ..., $\triangle h_{t-p}$, and the residuals are denoted by $\triangle\tilde{l}_t$, $\triangle\tilde{h}_t$ and $\tilde{h}_{t-1}$. By projection, the VEC form is simplified to (up to $o_p(1)$ terms):

$$\begin{cases} \triangle\tilde{l}_t = (\psi_{12}(1) + \beta\psi_{11}(1))\frac{c}{T}\tilde{h}_{t-1} + u_{1,t} \\ \triangle\tilde{h}_t = (\psi_{22}(1) + \beta\psi_{21}(1))\frac{c}{T}\tilde{h}_{t-1} + u_{2,t} \end{cases} \tag{3.2}$$

Let $c^* \equiv (\psi_{22}(1) + \beta\psi_{21}(1))c$, and use the condition $b_{12} = \frac{\psi_{12}(1)+\beta\psi_{11}(1)}{\psi_{22}(1)+\beta\psi_{21}(1)}$ implied by the long run restriction, we get:

$$\begin{cases} \triangle\tilde{l}_t = b_{12}\frac{c^*}{T}\tilde{h}_{t-1} + u_{1,t} \\ \triangle\tilde{h}_t = \frac{c^*}{T}\tilde{h}_{t-1} + u_{2,t} \end{cases} \tag{3.3}$$

Note that the condition $b_{12} = \frac{\psi_{12}(1)+\beta\psi_{11}(1)}{\psi_{22}(1)+\beta\psi_{21}(1)}$ holds only when $c \neq 0$, hence (3.3) is derived under $c \neq 0$. However, this simplified model remains correct when $c = 0$: $c = 0$ implies $c^* = 0$, hence $b_{12}c^* = 0$, no matter which expression of b_{12} is used. In other words, no matter whether or not c equals 0, it is valid to simplify (3.1) to (3.3).

Premultiplying the VEC form by B_0 yields the structural form below, with $B^{**}_{i-1} = B_0 A^{**}_{i-1}$:

$$B_0 \triangle \tilde{Y}_t = B_0(A^*(1) - I)\tilde{Y}_{t-1} + \sum_{i=1}^{p} B^{**}_{i-1} \triangle \tilde{Y}_{t-i} + \epsilon_t$$

Impose the long run restriction and write the structural form more explicitly:

$$\begin{cases} \triangle l_t = b_{12} \triangle h_t + \sum_{i=1}^{p} b^{**}_{i-1,11} \triangle l_{t-i} + \sum_{i=1}^{p} b^{**}_{i-1,12} \triangle h_{t-i} + \epsilon^z_t \\ \triangle h_t = b_{21} \triangle l_t + b^*_{22} h_{t-1} + \sum_{i=1}^{p} b^{**}_{i-1,21} \triangle l_{t-i} + \sum_{i=1}^{p} b^{**}_{i-1,22} \triangle h_{t-i} + \epsilon^d_t \end{cases}$$

where $b^*_{22} = [(\psi_{22}(1) + \beta\psi_{21}(1)) - b_{21}(\psi_{12}(1) + \beta\psi_{11}(1))]c/T$. After projecting out the lags, the structural form reduces to:

$$\begin{cases} \triangle \tilde{l}_t = b_{12} \triangle \tilde{h}_t + \epsilon^z_t \\ \triangle \tilde{h}_t = b_{21} \triangle \tilde{l}_t + b^*_{22} \tilde{h}_{t-1} + \epsilon^d_t \end{cases} \tag{3.4}$$

3.3 Tests

3.3.1 AR

For a given c, a confidence interval for b_{12} can be constructed by inverting AR test. Consider the auxiliary regression below for testing $H_0 : b_{12,0} = b_{12}$. Under H_0, $\theta = 0$:

$$\triangle \tilde{l}_t - b_{12,0} \triangle \tilde{h}_t = \theta \tilde{h}_{t-1} + \epsilon^z_t$$

The t_{AR} statistic is the t statistic for testing $\theta = 0$ in the auxiliary regression above:

$$t_{AR}(b_{12,0}) = \frac{(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1} \sum_{t=2}^{T} \tilde{h}_{t-1}(\triangle \tilde{l}_t - b_{12,0} \triangle \tilde{h}_t)}{(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \hat{\sigma}_\epsilon}$$

where $\hat{\sigma_\epsilon}^2$ is the sample variance of the OLS residual.

Theorem. *Under assumptions $i - iv$, and $H_0 : b_{12} = b_{12,0}$, t_{AR} has the asymptotic distribution that depends on c:*

$$t_{AR} \Longrightarrow \rho_{v\epsilon} \tau_c + \sqrt{1 - \rho_{v\epsilon}^2} z$$

where $\rho_{v\epsilon}$ is the long run correlation of v_t and ϵ_t^z, $v_t \equiv \sum_{i=1}^{p} a_{i-1,21}^{**} \triangle l_{t-i} + u_{2,t}$, $\tau_c \equiv \left(\int_0^1 J_c(s)^2 ds \right)^{-1/2} \int_0^1 J_c(s) dW(s)$, z is a standard normal variate.

Proof.

$$\begin{aligned} t_{AR}(b_{12,0}) &= \frac{(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1} \sum_{t=2}^{T} \tilde{h}_{t-1} \epsilon_t^z}{(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \hat{\sigma}_\epsilon} \\ &= \frac{(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} \epsilon_t^z}{\hat{\sigma}_\epsilon} \end{aligned}$$

Apply two results in Hansen (1995), with $a(1) \equiv \psi_{22}(1) + \beta \psi_{21}(1)$:

$$\frac{1}{T^2} \sum_{t=2}^{T} \tilde{h}_{t-1}^2 \Rightarrow a(1)^{-2} \sigma_v^2 \int_0^1 J^c(s)^2 ds$$

$$\frac{1}{T} \sum_{t=2}^{T} \tilde{h}_{t-1} \epsilon_t^z \Rightarrow a(1)^{-1} \sigma_v \sigma_\epsilon \left(\rho_{v\epsilon} \int_0^1 J^c(s) dW(s) + (1 - \rho_{v\epsilon}^2)^{1/2} \int_0^1 J^c(s) dW^\dagger(s) \right)$$

$$t_{AR}(b_{12,0}) \implies \left(\int_0^1 J_c(s)^2 ds\right)^{-1/2} \left(\rho_{v\epsilon}\int_0^1 J_c(s)dW(s) + \sqrt{1-\rho_{v\epsilon}^2}\int_0^1 J_c(s)dW^{\dagger}(s)\right)$$
$$\implies \rho_{v\epsilon}\tau_c + \sqrt{1-\rho_{v\epsilon}^2}z$$

□

3.3.2 Wald

A joint confidence set for (b_{12}, c^*) can be constructed by inverting Wald test.

Let $\phi_T(b_{12,0}, c_0^*) = \begin{pmatrix} (\sum_{t=2}^T \tilde{h}_{t-1}^2)^{-1} \sum_{t=2}^T \tilde{h}_{t-1}\triangle\tilde{l}_t - b_{12,0}\frac{c_0^*}{T} \\ (\sum_{t=2}^T \tilde{h}_{t-1}^2)^{-1} \sum_{t=2}^T \tilde{h}_{t-1}\triangle\tilde{h}_t - \frac{c_0^*}{T} \end{pmatrix}$. The Wald statistic is:

$$W(b_{12,0}, c_0^*) = \phi_T(b_{12,0}, c_0^*)' \left[\hat{\Sigma}_u^{-1} \sum_{t=2}^T \tilde{h}_{t-1}^2\right] \phi_T(b_{12,0}, c_0^*)$$

Theorem. *Under assumptions $i-iv$, and $H_0 : b_{12} = b_{12,0}, c^* = c_0^*$, $W(b_{12,0}, c_0^*)$ has the asymptotic distribution dependent on c:*

$$W(b_{12,0}, c_0^*) \Longrightarrow (\rho_{v2}\tau_c + (1-\rho_{v2}^2)^{1/2} z_2)^2 + (\rho_{v1-2}\tau_c + (1-\rho_{v1-2}^2)^{1/2} z_1)^2$$

where ρ_{v2} is the long run correlation of v_t and $u_{2,t}$, ρ_{v1-2} as the long run correlation of v_t with u_{1-2}, z_1, z_2 are independent standard normal variates, and u_{1-2} results from the decomposition of $u_{1,t} = \sigma_1(\frac{\rho_u}{\sigma_2}u_{2,t} + (1-\rho_u^2)^{1/2}u_{1-2})$, ρ_u is the correlation of $u_{1,t}$, $u_{2,t}$.

Proof. Decompose $u_{1,t}$:

$$u_{1,t} = \sigma_1(\frac{\rho_u}{\sigma_2}u_{2,t} + (1-\rho_u^2)^{1/2}u_{1-2})$$

where ρ_u is the correlation of $u_{1,t}$ and $u_{2,t}$, and u_{1-2} is uncorrelated with $u_{2,t}$: the subscript indicates that it comes from $u_{1,t}$ after excluding $u_{2,t}$; in addition, the variance of u_{1-2} is 1.

Let D_1 denote the convergence outcome of $(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1-2}$:

$$(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1-2} \Rightarrow D_1$$

where $D_1 = \rho_{v1-2}\tau_c + (1-\rho_{v1-2}^2)^{1/2} z_1$, and z_1 is a standard normal variate. Similarly:

$$(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{2,t} \Rightarrow \sigma_2 D_2$$

where $D_2 = \rho_{v2}\tau_c + (1-\rho_{v2}^2)^{1/2} z_2$, and z_2 is a standard normal variate independent of z_1.

With the above notations, rewrite $(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1,t}$ as:

$$\begin{aligned}
& \sigma_1 \left[(\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} (\frac{\rho_u}{\sigma_2} u_{2,t} + (1-\rho_u^2)^{1/2} u_{1-2}) \right] \\
= \; & \sigma_1 \left[\frac{\rho_u}{\sigma_2} (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{2,t} + (1-\rho_u^2)^{1/2} (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1-2} \right] \\
\Rightarrow \; & \sigma_1 \left[\rho_u D_2 + (1-\rho_u^2)^{1/2} D_1 \right]
\end{aligned}$$

Now the Wald statistic becomes:

$$\begin{pmatrix} (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1,t} \\ (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{2,t} \end{pmatrix}' {\Sigma_u}^{-1} \begin{pmatrix} (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{1,t} \\ (\sum_{t=2}^{T} \tilde{h}_{t-1}^2)^{-1/2} \sum_{t=2}^{T} \tilde{h}_{t-1} u_{2,t} \end{pmatrix}$$

$$\Rightarrow \quad \begin{pmatrix} \sigma_1 \left[\rho_u D_2 + (1-\rho_u^2)^{1/2} D_1\right] \\ \sigma_2 D_2 \end{pmatrix}' {\Sigma_u}^{-1} \begin{pmatrix} \sigma_1 \left[\rho_u D_2 + (1-\rho_u^2)^{1/2} D_1\right] \\ \sigma_2 D_2 \end{pmatrix}$$

$$\Rightarrow \quad \frac{\begin{pmatrix} \sigma_1 \left[\rho_u D_2 + (1-\rho_u^2)^{1/2} D_1\right] \\ \sigma_2 D_2 \end{pmatrix}' \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} \begin{pmatrix} \sigma_1 \left[\rho_u D_2 + (1-\rho_u^2)^{1/2} D_1\right] \\ \sigma_2 D_2 \end{pmatrix}}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}$$

$$\Rightarrow \quad D_1^2 + D_2^2$$

□

We are ultimately most interested in the IRF $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$: the effect of a positive technology shock on the hours worked. Once we get b_{12} by inverting AR or Wald, we derive $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$ in the following manner.

Given b_{12}, there is a correspondent b_{21}: $b_{21} = \frac{b_{12}\sigma_{22} - \sigma_{12}}{b_{12}\sigma_{12} - \sigma_{11}}$. This relation holds because the covariance matrix $E(\epsilon_t \epsilon_t')$ of the structural shocks is diagonal, and $B_0 E(u_t u_t') B_0' = E(\epsilon_t \epsilon_t')$. Consequently, $B_0 = \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}$ can be recovered.

Combining B_0 with $A_1, ..., A_{p+1}$, which are estimated by VAR$(p+1)$ of $Y_t =$

$\begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix}$, the IRF $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$ can be derived through the companion matrix F:

$$F = \begin{pmatrix} A_1 & A_2 & ... & A_{p+1} \\ I & & & \\ & ... & & \\ & & I & \end{pmatrix}$$

Take the j^{th} power of F, and denote the up left 2 by 2 matrix of F^j by $\tilde{F}^j$, then $\frac{\partial Y_{t+j}}{\partial \epsilon_t{}'} = \tilde{F}^j B_0^{-1}$, and $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$ is the $(2,1)$ element of $\frac{\partial Y_{t+j}}{\partial \epsilon_t{}'}$.

As an example, when $j = 0$, the contemporaneous effect of $\frac{\partial h_t}{\partial \epsilon_t^z}$ is:

$$\begin{aligned} \frac{\partial h_t}{\partial \epsilon_t^z} &= \frac{b_{21}}{1 - b_{12}b_{21}} \\ &= \frac{b_{12}\sigma_{22} - \sigma_{12}}{2b_{12}\sigma_{12} - \sigma_{11} - b_{12}^2\sigma_{22}} \end{aligned}$$

Once a robust confidence interval of b_{12} is available, we derive the confidence interval for $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$ by using the point estimates of $A_1, ..., A_{p+1}, \Sigma_u$, and all possible values of b_{12} in its confidence interval: the maximum/minimum value of $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$ becomes its upper/lower bound, respectively.

3.4 Application

The same data as in Christiano *et al.* (2003) and Pesavento and Rossi (2005) is used in our empirical application: the quarterly data of labor productivity and the hours worked (denoted by l_t, h_t respectively in this chapter) between $1948Q1-2001Q4$ are from the DRI Economics Database; labor productivity is measured by the output per hour in the business sector, while hours is measured by the hours worked in the business sector divided by the population; both labor productivity and hours are in natural logarithm. See Christiano *et al.* (2003) for further details.

Although Christiano *et al.* (2003) and Pesavento and Rossi (2005) use the same dataset, their conclusions contradict each other: Christiano *et al.* (2003) argue that the level specification of hours should be chosen over the difference specification, and report that a positive technology shock will drive hours up, while Pesavento and Rossi (2005) employ an agnostic procedure which works for both the level and difference specifications of hours, and report that a positive productivity shock drives hours down.

The recent work by Gospodinov *et al.* (2009), Gospodinov (2010) suggest that it is appropriate to use the model of (3.1) to investigate the impact of technology shocks on hours, and we adopt the robust approach for (3.1) to construct the confidence interval for the response of hours to technology shocks. Since we are using the same dataset, it will be interesting to compare our results with Christiano *et al.* (2003), Pesavento and Rossi (2005), etc.

As a starting point, we consider the two specifications of hours, in levels and in first differences. The conventional identification procedure (see, e.g., Gali (1999)) under the restriction that only technology shocks have a long run effect on labor pro-

ductivity is applied for the two specifications, respectively: following Christiano *et al.* (2003), a VAR model with a drift and four lags is employed, where l_t is specified in first differences, h_t is specified either in levels or in first differences.

The response of the hours worked to the positive technology shocks is plotted in Figure 3.1, and it is consistent with the current debate: at short horizons, the effect of a positive technology shock on the hours worked is found positive under the level specification of hours, and this effect is negative under the difference specification. The 95% confidence bounds are constructed by the bootstrap.

For future use, we estimate the following matrices: (i) the coefficient matrix A_1, A_2, A_3, A_4 in the VAR(4) model of $\begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix}$; (ii) the covariance matrix Σ_u of the reduced error u_t; (iii) the long run matrix $\Psi(1)$.

i. $\hat{A}_1, \hat{A}_2, \hat{A}_3, \hat{A}_4$: estimate a VAR(4) model with a drift for $\begin{pmatrix} \triangle l_t \\ h_t \end{pmatrix}$, $\hat{A}_1, \hat{A}_2, \hat{A}_3, \hat{A}_4$ are the OLS estimates of the coefficients before the first, second, third, fourth lag, respectively; the companion matrix $F = \begin{pmatrix} A_1 & A_2 & A_3 & A_4 \\ I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \end{pmatrix}$ is thus

estimated by:

$$\hat{F} = \begin{pmatrix} -0.1008 & 0.0698 & 0.0317 & -0.2719 & -0.0616 & 0.0858 & -0.0286 & 0.1344 \\ 0.1611 & 1.4193 & 0.2083 & -0.4265 & 0.1303 & 0.0131 & 0.0503 & -0.0329 \\ 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 & 0 & 0 \end{pmatrix}$$

ii. $\hat{\Sigma}_u$: save the residuals of the VAR(4) model above, and $\hat{\Sigma}_u = \begin{pmatrix} 0.8707 & 0.0015 \\ 0.0015 & 0.5390 \end{pmatrix}$ is the sample covariance matrix of the residuals.

iii. $\hat{\Psi}(1)$: estimate a VAR(3) model with a drift for $\begin{pmatrix} \triangle l_t \\ \triangle h_t \end{pmatrix}$, $\hat{\Psi}_1, \hat{\Psi}_2, \hat{\Psi}_3$ are the OLS estimates of the coefficients before the first, second, third lag, respectively, and:

$$\hat{\Psi}(1) = 1 - \hat{\Psi}_1 - \hat{\Psi}_2 - \hat{\Psi}_3 = \begin{pmatrix} 1.0895 & 0.2986 \\ -0.4372 & 0.4970 \end{pmatrix}$$

Using the same dataset, Pesavento and Rossi (2005) report that the CADF (Covariate Augmented Dickey-Fuller) statistic is around -3.072, and the 95% confidence interval of ρ is $(0.897, 0.997)$. When $T = 216$ in the application, it implies that the 95% C.I. of c is: $(c^l, c^u) = (-22.2480, -0.6480)$.

Take the following steps to invert AR and construct C.I. for the IRF:

1. For each $c \in (c^l, c^u)$, invert t_{AR} for a C.I. of b_{12};

2. Bonferroni: take the union of C.I.'s of b_{12} from above;

3. For each b_{12} in its C.I., compute $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$;

4. Upper/lower bounds of C.I. come from max/min of $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$.

Similarly, take the following steps to invert Wald test and construct C.I. for the IRF:

1. For each $(b_{12,0}, c_0^*)$, compute the Wald statistic, and save $(b_{12,0}, c_0^*)$ if $W(b_{12,0}, c_0^*)$ does not exceed the critical value;

2. Construct the C.I. of $b_{12,0}$ by projection;

3. For each b_{12} in its C.I., compute $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$;

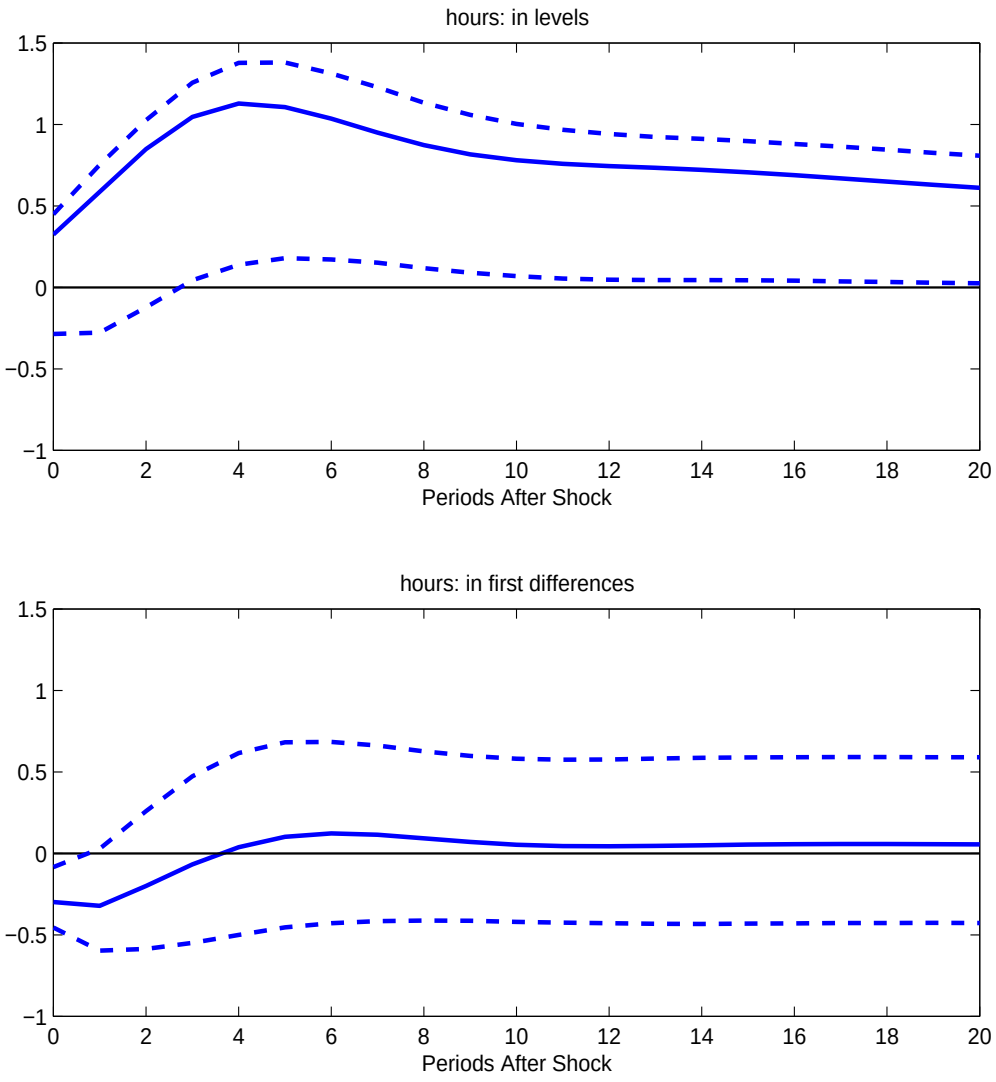
4. Upper/lower bounds of C.I. come from max/min of $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$.

The response of the hours worked to the positive technology shocks by AR and Wald is plotted in Figure 3.2. It is found that the IRF lies mostly in the positive region at short horizons, although at 95%, we can not reject that the contemporaneous effect is negative.

3.5 Conclusion

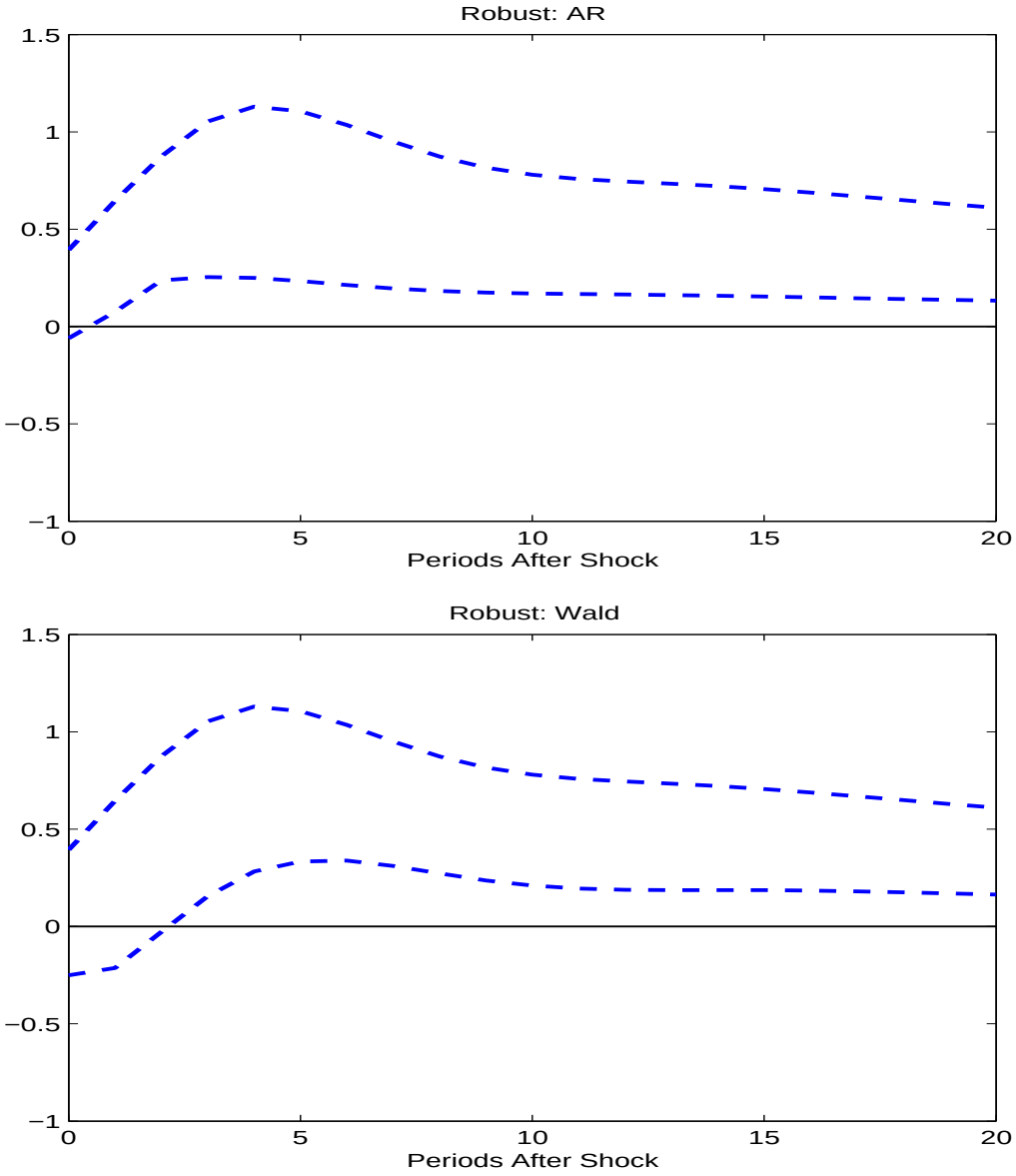
We contribute to the current debate on whether a positive technology shock will bring the hours worked up or down by providing new robust results: this effect on hours is found most positive at short horizons, although we do not rule out the possibility that the contemporaneous effect is negative at 95%.

Figure 3.1: IRF under level and first-difference



Notes: this figure reports the IRF $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$, i.e. the effect of positive technology shocks on the hours worked, and the IRF is constructed under the long run restriction by the same approach as in Christiano *et al.* (2003). The solid line is the point estimate, with h_t in levels in the upper panel, and h_t in first differences in the low panel; the dashed line is the 95% upper and lower bounds by the bootstrap.

Figure 3.2: IRF bounds by AR and Wald



Notes: this figure reports the 95% bounds of the IRF $\frac{\partial h_{t+j}}{\partial \epsilon_t^z}$, i.e. the effect of positive technology shocks on the hours worked, and the bounds are constructed by the AR and Wald approach proposed in this chapter.

BIBLIOGRAPHY

Acharya, V. and Pedersen, L. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, **77** (2), 375–410.

Anderson, T. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, pp. 327–351.

— and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, pp. 46–63.

Angrist, J. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, pp. 979–1014.

Bravo, F., Escanciano, J. and Otsu, T. (2009). Testing for Identification in GMM under Conditional Moment Restrictions, working Paper.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling, Aspects of labour market behaviour: essays in honour of John Vanderkamp. *ed. L.N. Christofides, E.K. Grant, and R. Swidinsky*.

Christiano, L., Eichenbaum, M. and Vigfusson, R. (2003). What happens after a technology shock? *NBER working paper*.

Cochrane, J. (2001). *Asset pricing*. Princeton University Press.

Cragg, J. and Donald, S. (1996). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association*, pp. 1301–1309.

David, H. and Nagaraja, H. (2003). *Order statistics*. Wiley-Interscience.

Davidson, R. and MacKinnon, J. (2010). Wild bootstrap tests for IV regression. *Journal of Business and Economic Statistics*, **28** (1), 128–144.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, pp. 1–26.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of finance*, pp. 427–465.

— and — (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, **33** (1), 3–56.

Francis, N. and Ramey, V. (2005). Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. *Journal of Monetary Economics*, **52** (8), 1379–1399.

Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, **12** (3), 827–842.

Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *The American Economic Review*, **89** (1), 249–271.

Gospodinov, N. (2010). Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions. *Journal of Business and Economic Statistics*, **28** (1), 1–12.

—, Maynard, A. and Pesavento, E. (2009). Sensitivity of Impulse Responses to Small Low Frequency Co-Movements: Reconciling the Evidence on the Effects of Technology Shocks. *Cahiers de recherche*.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, **70** (1), 163–189.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer Verlag.

— and Horowitz, J. (1996). Bootstrap critical values for tests based on generalized method of moments estimators. *Econometrica*, **64** (4), 891–916.

Hansen, B. (1995). Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econometric Theory*, **11** (05), 1148–1171.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

—, Heaton, J. and Li, N. (2008). Consumption strikes back? measuring long-run risk. *Journal of Political Economy*, **116** (2), 260–302.

Horowitz, J. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, **61** (2), 395–411.

— (2001). The Bootstrap. Handbook of Econometrics, Vol. 5.

Inoue, A. and Rossi, B. (2008). Testing for weak identification in possibly nonlinear models.

Jagannathan, R. and Wang, Z. (1996). The conditional capm and the cross-section of expected returns. *Journal of Finance*, pp. 3–53.

Kan, R. and Zhang, C. (1999). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, pp. 203–235.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, pp. 1781–1803.

— (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, pp. 1103–1123.

— (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, **149** (2), 149–173.

— and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133** (1), 97–126.

Lettau, M. and Ludvigson, S. (2001). Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, **109** (6), 1238–1287.

Lewellen, J., Nagel, S. and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, **96** (2), 175–194.

Li, Q., Vassalou, M. and Xing, Y. (2006). Sector investment growth rates and the cross section of equity returns. *The Journal of Business*, **79** (3), 1637–1665.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, pp. 13–37.

Lustig, H. and Van Nieuwerburgh, S. (2005). Housing collateral, consumption insurance, and risk premia: An empirical perspective. *Journal of Finance*, pp. 1167–1219.

Moreira, M. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, pp. 1027–1048.

—, Porter, J. and Suarez, G. (2009). Bootstrap validity for the score test when instruments may be weak. *Journal of Econometrics*, **149** (1), 52–64.

Parker, J. and Julliard, C. (2005). Consumption risk and the cross section of expected returns. *Journal of Political Economy*, **113** (1), 185–222.

Pesavento, E. and Rossi, B. (2005). Do technology shocks drive hours up or down? A little evidence from an agnostic procedure. *Macroeconomic Dynamics*, **9** (04), 478–488.

Piazzesi, M., Schneider, M. and Tuzel, S. (2007). Housing, consumption and asset pricing. *Journal of Financial Economics*, **83** (3), 531–569.

Rebelo, S. (2005). Real business cycle models: past, present and future. *Scandinavian Journal of Economics*, **107** (2), 217–238.

Rothenberg, T. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, **2**, 881–935.

Santos, T. and Veronesi, P. (2006). Labor income and predictable stock returns. *Review of Financial Studies*, **19** (1), 1–44.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, pp. 1–33.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of finance*, pp. 425–442.

Shea, J. (1999). What do technology shocks do? *NBER Macroeconomics Annual*.

Silverman, B. (1998). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, pp. 557–586.

Stock, J. and Wright, J. (2000). GMM with weak identification. *Econometrica*, pp. 1055–1096.

—, — and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, **20** (4), 518–529.

— and Yogo, M. (2005). Testing for weak instruments in linear IV regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. *ed. DW Andrews and JH Stock*, pp. 80–108.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, **48** (4), 817–838.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. MIT press.

Wright, J. (2002). Testing the Null of Identification in GMM. *International Finance Discussion Papers*, **732**.

— (2003). Detecting lack of identification in GMM. *Econometric Theory*, **19** (02), 322–330.

Yogo, M. (2006). A consumption-based explanation of expected stock returns. *The Journal of Finance*, **61** (2), 539–580.