Prices, Holdings, and Learning in Financial Markets. Experiments and Methodology

Thesis by

Debrah Meloso

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2007
(Defended May 15, 2007)

Pa Otila, mi mama

Acknowledgements

Thank you Peter, Bill, and John! Peter taught me that science is one, amazing, never-ending pursuit. It was also in Peter's office that I learned a big part of all the Finance and Econometrics I know. Bill has worked tirelessly to eliminate my student mentality and replace it with the courage, imagination, self-criticism, and discipline of an independent scientist and economist. I must make a special note of Bill's dedication to being my mentor and adviser. He has always been there, at moments that were clearly important, and also those that only seemed important to me. I received two very big gifts from John during my time at Caltech: John understood better than me that I wished to be a good economist, and he always told me to work on research I loved.

I have had numerous other mentors and advisers. I must mention Colin and his enthusiasm and support for my most daring projects – and his opinion that they were not daring enough. At Caltech I met and benefitted from interacting with Ken Binmore, Preston McAfee, and Jaksa Cvitanic. All faculty members have contributed to make me a better person. During my masters at UAB, Clara Ponsati took upon herself the task of teaching me from scratch what research is, and tried really hard to imbue me with her enthusiasm for theory. I thank her for believing in me and for remaining a very dear mentor. My peers at UAB are still great friends and coauthors. I cannot imagine a more stimulating intellectual environment! At Caltech I met Andrea Galeotti (Mimmo), who helped me bring to reality my romantic idea of *doing science over a beer*. I find it very hard to do science in this way when Mimmo is not around, except, of course, with the "girls." They broke my belief that peers in the USA were detached and unromantic about science. Jernej has, at one moment or the other, assumed many of the roles I described until now. He has been my peer at UAB and Caltech, the type of person I can do science with over a beer, and sometimes a mentor. Thank you!

My family (Meme, my aunts, uncles and cousins) has been great. For years it has been an absolute mystery to them what I am doing and where I am going, but they kept believing it was something to be proud of. Aren't they wonderful? My mom was the main engine of this energy. My in-laws, Alenka, Hela and Tine, were also always supportive and stimulating. Some crucial friends, because they were there at crucial moments or because they are crucial to me at all times, include Bill and Linda, Mimmo, Paul, Lucia, Martin, Sera, Maggie, Meghana, Tomomi, Lissethe, Laura and Carmelo, and Andre. Thank you all!

Abstract

This thesis is a compilation of three essays that bridge the theoretical and empirical study of financial markets. The subjects of study in the three main chapters are (i) equilibrium models of asset prices and asset holdings and trade; (ii) limited computational capacity and its interaction with asset prices and trades.

In chapter 1 (joint with Peter Bossaerts) we show that statistical improvements can be made on a traditional test of portfolio "efficiency." Testing portfolio efficiency is used in the practice of investment decisions as well as to test theoretical models of asset prices (CAPM and multifactor models). We propose a parametric family of tests of the efficiency of a portfolio in a market with a risk-free asset. All tests in the family compare the mean-variance ratio of the tested portfolio (*benchmark*) with that of a different portfolio (*reference*). We show that the power of a test in our proposed family depends on the correlation between the benchmark and the reference portfolio. This provides a way to improve the power of efficiency tests for a given sample, by choosing the appropriate test in this family.

Chapter 2 (joint with Peter Bossaerts and William Zame), is a test of the theory of dynamically complete markets. In this work we compare prices and portfolio choices in complete and incomplete experimental financial markets. The incomplete-markets treatment differs from the complete-markets one in that we close one market, and announce, halfway through trading, which of three states will not occur. We find prices and allocations to be analogous across the two treatments, as predicted by theory. In particular, subjects' additional trading in the incomplete-markets treatment is such that the final allocations become indistinguishable from the complete-markets treatment. The results show that participants form rational expectations about retrade prices, which is a very strong finding.

Chapter 3 (joint work with Peter Bossaerts and Jernej Copic) moves away from existing theoretical paradigms. It explores the implications of analyzing intellectual discovery as the solution of a nonincremental problem, outside the reach of traditional models of learning with updating. The experiment sets up a situation that is non-incremental and where Bayesian updating is not a sensible model. In this framework we find that communication is possible, and that a primitive code is good enough to achieve intellectual discovery, not discourage it.

Contents

Chapter 1

Portfolio Correlation and the Power of Portfolio Efficiency Tests

Abstract

We propose a parametric family of tests of the mean-variance efficiency of a portfolio in a market with a risk-free asset. All tests in the family compare the mean-variance ratio of the tested portfolio (the *benchmark* portfolio) with the same ratio for a different portfolio, called the *reference* portfolio. The Gibbons-Ross-Shanken test belongs to this family, and the reference in this case is the ex-post tangency portfolio of the market. We show that the power of a test in our proposed family depends on the correlation between the benchmark and the reference portfolio. This correlation, and thus, the power of the test, can be manipulated by changing the value of the parameter that spans the family. In particular, for a given sample, a power maximizing test can be easily found in the family we propose. This power-maximizing test will generically not be the Gibbons-Ross-Shanken test.

1.1 Introduction

The mean and the variance of a portfolio are the most commonly used and reported measures of its goodness. Mean-variance efficiency is a minimal criterion to sort portfolios along these two dimensions of performance, and is therefore widely used and reported. On the other hand, the mean-variance efficiency of the *market* portfolio is one of the main results of the CAPM model, adding to the reasons why the measurement of mean-variance efficiency is relevant.[1]

If a portfolio is mean-variance efficient, the mean returns of all assets can be represented as coming from two sources, one of them being the mean-variance efficient portfolio, and the other being an asset with fixed returns. In other words, mean excess returns (excess over the return of an asset with fixed return) of all assets are a linear transformation of the mean excess return of a mean-variance efficient portfolio. If the distribution of returns of all assets is known, the above result can be readily used to establish what portfolios are mean-variance efficient. In reality, the distribution of returns is unknown and must be recovered from a sample of realizations. The statistical equivalent to verifying the linear transformation result is to establish whether the *Jensen* alphas are statistically different from zero. This question spans a class of tests used to establish mean-variance efficiency of an arbitrary portfolio, to test the CAPM if the market portfolio is used, and to test other, non-equilibrium, asset pricing models, like the multi-factor models. This is why much of what is claimed about the empirical success or failure of asset pricing models, as well as about portfolio performance in general, depends on the statistical properties of this class of tests.

We propose a family of tests that lies in this class. It is based on the GRS test proposed in Gibbons, Ross, and Shanken [1989], and comprises this test as a special case. Under the assumption that a risk-free asset exists, the GRS test is nowadays the standard test of mean-variance efficiency, because of its simple way of dealing with the twofold statistical problem of multiple hypothesis testing and serial correlation. It has, however, been subjected to the criticism of having low statistical power.[2] This is the issue we deal with. On one hand, we point at the source of the power properties of the family that we propose (including the GRS test), and the Gibbons test (see Gibbons [1982]).[3] On the other hand, we reduce the problem of finding a powerful test - within the family we propose - to that of changing a single parameter.

The idea behind the possibility of improving upon the power of the GRS test is very simple. Suppose a statistician is faced with the task of finding out whether the means of two random variables, X and Y, are equal. To do this, she is endowed with two tools. One is a finite sample taken from the true joint distribution of X and Y. The other tool is the power of choosing the correlation of X and Y. It is clear that, given such power, she will choose maximal correlation, since this minimizes the variance of the statistical inference she can make from the sample. Statisticians rarely have this power. However, the econometrician that is trying to assess the mean-variance efficiency of a portfolio implicitly has this power. The family of test

[1] Mean-variance efficiency of the market portfolio can also be derived as an "approximate" result in equilibrium pricing models with assumptions on preferences or return distributions that differ from those in the CAPM.

[2] The power of a test is the probability of rejecting the null hypothesis when it is false.

[3] The Gibbons test is the equivalent of the GRS test in the absence of a risk-free asset. If used in the presence of a risk-free asset, it is in principle a weaker test than the GRS. Nonetheless, the Gibbons test has in occasions rejected the null hypothesis where the GRS has not. This fact lies at the heart of the claim that the GRS has low power.

statistics that we propose is based on the comparison between two portfolios (like the comparison between X and Y that the statistician had to make). The GRS test fixes this comparison to be between the sample mean-variance efficient portfolio and the portfolio whose mean variance efficiency is being tested. We claim that the latter need not be compared to the sample mean-variance efficient portfolio in order to asses its efficiency. In fact, the family of tests we propose is given by a parameter identifying other candidates for comparison. Picking a different candidate is equivalent to choosing a different covariance for the random variables being compared. This is how statistical power can be increased.

Of course, the econometrician does not know the true covariance of portfolios, that can lead to improved power. An estimate must be used instead. We use this estimate to find the power maximizing test within the family we propose, for a sample of CRSP data on portfolio returns. We show that power improvements over the GRS test can be very significant.

In the following section we introduce notation and define mean-variance efficiency. In section 3 we propose a family of mean-variance efficiency tests and give its statistical properties. In section 4 we develop the intuition that tests of mean-variance efficiency are based on the comparison between two portfolios, and give the result that the power of the test depends on the true correlation between these portfolios. We analyze the power of the proposed family of tests in section 5. In section 6 we give numerical examples of applications of our proposed power improving tests. We conclude in section 7, suggesting extensions of this work, and a possible answer to the puzzle mentioned in footnote 3.

1.2 Problem Statement

Throughout this paper we will use *efficiency* to refer to mean-variance efficiency of a portfolio. We proceed to define this latter notion, after giving the necessary setup and notation.

Let $\{\bar{R}, \Omega\}$ denote a market. For our purpose a market is fully defined by mean and covariance of asset returns. Risky assets have random returns denoted $R = (R_1, \ldots, R_N)$. The vector of mean returns of risky assets is denoted $\bar{R} = \left(\bar{R}_1, \ldots, \bar{R}_N\right)$, and Ω is the $N \times N$ matrix of asset return covariances. We will mostly consider markets where there is a risk-free asset, with fixed, known return R_F. For clarity, we will use $\{\bar{R}, \Omega\}$ whenever referring to a market with no risk-free asset, and $\{\bar{R}, \Omega, R_F\}$ whenever referring to a market with a risk-free asset.

A portfolio, p, is a combination of assets. Let $w_p = (w_{p1}, \ldots, w_{pN})$ denote the relative holdings of each risky asset that makes up portfolio p. Each weight is the fraction of total wealth invested in the corresponding asset. The residual, $(1 - w_p'\mathbf{1})$, gives the relative holding of the risk-free security. The mean return of portfolio p is $\bar{R}_p = w_p'\bar{R} + (1 - w_p'\mathbf{1})R_F$, and its variance is $\sigma_p^2 = w_p'\Omega w_p$. The covariance between portfolio p and all risky assets is the vector $(\sigma_{p1}, \ldots, \sigma_{pN}) = w_p'\Omega$, and the covariance of two portfolios p and q is $\sigma_{pq} = w_p'\Omega w_q$.

The Sharpe ratio of portfolio p is defined to be

$$SR^p = \frac{\bar{R}_p - R_F}{\sigma_p}.$$

The Sharpe ratio of a portfolio is its measure of efficiency. Given $\{\bar{R}, \Omega, R_F\}$, a portfolio is *efficient* if it has maximal Sharpe ratio. Alternatively, we say that *portfolio e is efficient* if w_e solves the following program:

$$\min_{w} w'\Omega w \qquad (1.1)$$
$$\text{s.t.} \quad \bar{R}_e = w'\bar{R} + (1 - w'\mathbf{1})R_F$$

for a target mean return $\bar{R}_e$.

It is clear from (1.1) that there are many efficient portfolios, depending on the chosen target mean return, $\bar{R}_e$. All efficient portfolios in $\{\bar{R}, \Omega, R_F\}$ have the same Sharpe ratio, which clearly means that they are located on a line in mean-variance space. In other words, if program (1.1) is solved for all R_e, the solution is a linear relation between the mean and standard deviation of efficient portfolios, with slope equal to the maximal Sharpe ratio. Specifically, the solutions to program (1.1) define the following *efficient frontier* for market $\{\bar{R}, \Omega, R_F\}$:

$$\bar{R}_e = R_F + \left(\sqrt{(\bar{R} - R_F\mathbf{1})'\Omega^{-1}(\bar{R} - R_F\mathbf{1})}\right)\sigma_e,$$

and the maximal Sharpe ratio in market $\{\bar{R}, \Omega, R_F\}$, is the slope in the above expression,

$$SR^* = SR^e = \frac{\bar{R}_e - R_F}{\sigma_e} = \sqrt{\left(\bar{R} - R_F\mathbf{1}\right)'\Omega^{-1}\left(\bar{R} - R_F\mathbf{1}\right)}.$$

The efficient frontier in market $\{\bar{R}, \Omega, R_F\}$ is affine because of the presence of the risk-free asset. One can linearly trade off mean return and standard deviation by changing the relative holding of the risk-free asset with respect to a fixed portfolio of risky assets. This fixed portfolio of risky assets is the *tangency portfolio* corresponding to R_F, which will be denoted by t, and has weights $w_t = \frac{\Omega^{-1}(\bar{R} - R_F\mathbf{1})}{\mathbf{1}'\Omega^{-1}(\bar{R} - R_F\mathbf{1})}$.

The tangency portfolio for market $\{\bar{R}, \Omega, R_F\}$, is efficient in this market as well as in market $\{\bar{R}, \Omega\}$, where there is no risk-free asset. It lies at the tangency of the efficient frontier for market $\{\bar{R}, \Omega, R_F\}$ and that of market $\{\bar{R}, \Omega\}$. The latter is not a line in mean-standard deviation space. It is a parabola that lies within the efficient frontier for the market with a risk-free asset, since more combinations of assets can be achieved in the larger market, $\{\bar{R}, \Omega, R_F\}$. Figure 1.1 shows an example of efficient frontier for a market without risk-free asset, the efficient frontier once a risk-free asset is added, and the tangency portfolio for $\{\bar{R}, \Omega, R_F\}$.

An important property of an efficient portfolio is that, for all assets in the market, the excess return (difference of the asset's return and the return of the risk-free asset) of an asset is a linear transformation of the excess return of the efficient portfolio. This property lies at the heart of the family of tests that we will propose, as well as the GRS test of Gibbons, Ross, and Shanken [1989]. Thus, for $n \in \{1, \ldots, N\}$ it holds that

$$\bar{R}_n - R_F = \frac{\sigma_{en}}{\sigma_e^2}(\bar{R}_e - R_F). \qquad (1.2)$$

For the sake of completeness, we include one derivation of the above expression. Define a portfolio ϕ

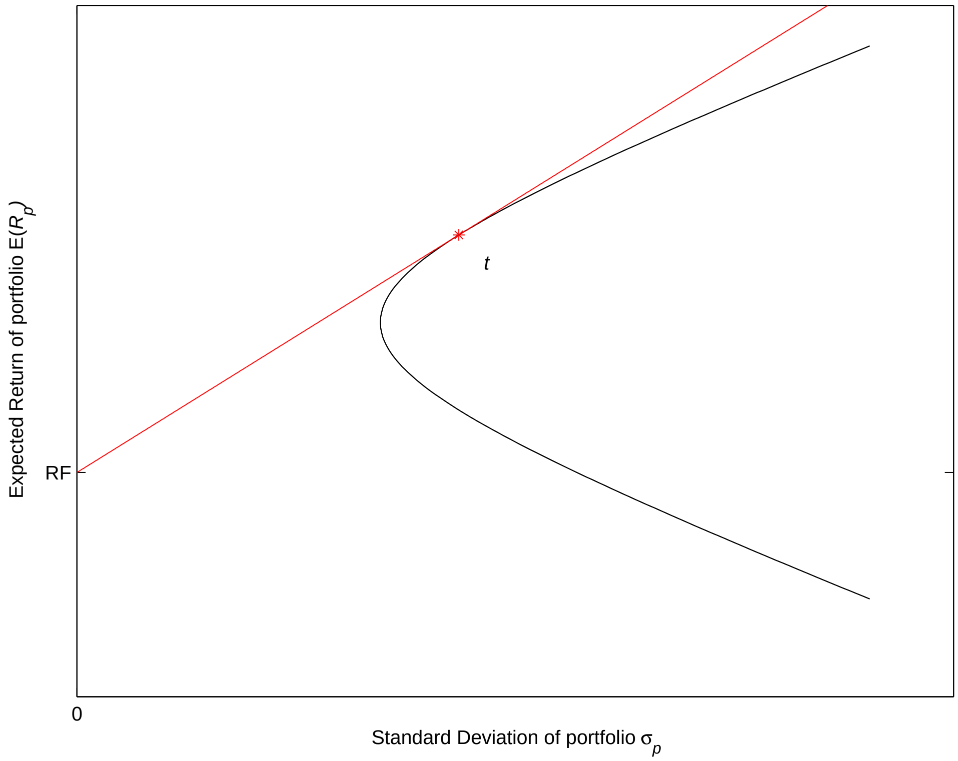


Figure 1.1. Efficient frontier for a market with a risk-free asset (straight line), and for a market without a risk-free asset (curve).

with the following mean and variance of returns:

$$\bar{R}_\phi = \phi\bar{R}_e + (1-\phi)\,\bar{R}_p,$$
$$\sigma_\phi = \left[\phi^2\sigma_e^2 + (1-\phi)^2\,\sigma_p^2 + 2\phi\,(1-\phi)\,\sigma_{ep}\right]^{\frac{1}{2}},$$

where p is an arbitrary portfolio. The Sharpe ratio of the efficient portfolio can be found as the derivative of $\bar{R}_\phi$ with respect to σ_ϕ, evaluated at $\phi = 1$. Thus,

$$\left.\frac{\partial\bar{R}_\phi}{\partial\sigma_\phi}\right|_{\phi=1} = -\left.\frac{\frac{\partial\bar{R}_\phi}{\partial\phi}}{\frac{\partial\sigma_\phi}{\partial\phi}}\right|_{\phi=1} = \frac{\bar{R}_e - R_F}{\sigma_e}.$$

Solving for the above equation gives the result in (1.2).

The question we deal with in this paper is whether a given portfolio, called the benchmark, is efficient in the market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$, where $\bar{R}_{+b} = \left(\bar{R}_1, \ldots, \bar{R}_N, \bar{R}_b\right)$ is an $(N+1)\times 1$ vector of mean returns of assets, where the $N+1$th asset is portfolio b, and Ω_{+b} is the variance-covariance matrix of the vector of returns R_{+b}. The question of the efficiency of the benchmark is clearly equivalent to asking whether it is true that

$$\bar{R}_n - R_F = \frac{\sigma_{bn}}{\sigma_b^2}\left(\bar{R}_b - R_F\right)$$

for $n \in \{1, \ldots, N\}$.

1.3 A Translation Family of Tests

Define $R_\gamma = R_F - \gamma$, where γ is an arbitrary real number. Efficiency of portfolio b can be restated to be the condition that

$$\bar{R}_n - R_\gamma = \frac{\sigma_{bn}}{\sigma_b^2}\left(\bar{R}_b - R_\gamma\right) + \left(1 - \frac{\sigma_{bn}}{\sigma_b^2}\right)\gamma \tag{1.3}$$

hold for $n \in \{1, \ldots, N\}$.

A regression-based empirical test of the efficiency of a benchmark portfolio, b, can be defined for every γ, thus defining a parametric family of efficiency tests.

1.3.1 Regression and Joint Null Hypothesis Test

In market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$, let $\begin{bmatrix} R^T & R_F^T \end{bmatrix}$ be a sample of size T taken from the true distribution of asset returns, and let $R_\gamma^T = R_F^T - \gamma \mathbf{1}_T$, be the sample of R_F translated by γ. Fix γ and define the following projection of risky assets excess returns (in excess of R_γ) on the excess return of the benchmark portfolio:

$$\begin{aligned} &R_{n\tau} - R_{\gamma\tau} = \alpha_n\left(\gamma\right) + \beta_{\gamma n}(R_{b\tau} - R_{\gamma\tau}) + \epsilon_{n\tau}, \\ &E[\epsilon_\tau] = \mathbf{0}, \quad E[\epsilon_\tau(R_{b\tau} - R_{F\tau})] = E[\epsilon_\tau(R_{b\tau} - R_{\gamma\tau})] = \mathbf{0}, \\ &n \in \{1, \ldots, N\}, \quad \tau \in \{1, \ldots, T\}, \end{aligned} \tag{1.4}$$

where τ identifies the sample point or time, and $\epsilon_\tau = (\epsilon_{1\tau}, \ldots, \epsilon_{N\tau})$ is a vector of errors, assumed to be independent across τ.

If b is (mean-variance) efficient, equation (1.3) holds for all assets. The null hypothesis is thus a multivariate hypothesis on the value of $\alpha\left(\gamma\right) = \left(\alpha_1\left(\gamma\right), \ldots, \alpha_N\left(\gamma\right)\right)$ and $\beta_\gamma = (\beta_{\gamma 1}, \ldots, \beta_{\gamma N})$. By construction, in the above projection, $\beta_{\gamma n} = \frac{\sigma_{bn}}{\sigma_b^2}$. Thus, the null hypothesis reduces to the hypothesis on the value of $\alpha\left(\gamma\right)$ given below:

$$\begin{gathered} H_0: \quad \alpha\left(\gamma\right) = \left(1 - \beta_{\gamma n}\right)\gamma \\ \forall n \in \{1, \ldots, N\}. \end{gathered}$$

Let $\hat{\alpha}\left(\gamma\right)$ and $\hat{\beta}_\gamma$ be the least-squares estimators of $\alpha\left(\gamma\right)$ and β_γ. Asymptotically, or under the additional assumption that ϵ_τ are normally distributed, $\hat{\alpha}\left(\gamma\right)$ satisfies

$$\hat{\alpha}\left(\gamma\right) \sim N\left(\alpha\left(\gamma\right), \frac{1}{T}\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]\Sigma\right),$$

where

$$\hat{\mu}_{b\gamma} = \frac{1}{T}\sum_{\tau=1}^{T}\left(R_{b\tau} - R_{\gamma\tau}\right), \text{ and}$$
$$\hat{\sigma}_b^2 = \frac{1}{T}\sum_{\tau=1}^{T}(R_{b\tau} - R_{F\tau} - \hat{\mu}_b)^2.$$

For expositional ease we will assume normality of the white noise error terms,

$$\epsilon_\tau \sim N(\mathbf{0}, \Sigma),$$

where Σ is the $N \times N$ matrix of covariances of $\epsilon_{n\tau}$ over n.

Given the least-squares regression estimates, $\hat{\alpha}(\gamma)$, there are several ways to test the joint null hypothesis, $\alpha_n(\gamma) = 0$ for all n. We follow Gibbons, Ross, and Shanken [1989], and propose the following test:

Definition 1 (γ-GRS test)**.** *We define the γ-GRS test of the efficiency of a portfolio with the estimator*

$$J(\gamma) = T\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]^{-1}\left[\hat{\alpha}(\gamma)\right]'\Sigma^{-1}\left[\hat{\alpha}(\gamma)\right],$$

and the null hypothesis

$$H_0: \quad \alpha(\gamma) = (\mathbf{1} - \beta_\gamma)\gamma.$$

Under our projection assumptions, $J(\gamma)$ has a non-central chi-square distribution with N degrees of freedom and non-centrality parameter $\lambda(\gamma)$. That is,

$$\begin{aligned} &J(\gamma) \sim \chi_N^2(\lambda(\gamma)), \\ &\text{where } \lambda(\gamma) = T\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]^{-1}\alpha(\gamma)'\Sigma^{-1}\alpha(\gamma). \end{aligned} \tag{1.5}$$

When Σ is unknown, it can be replaced with a consistent estimator $\hat{\Sigma}$. In that case, the finite-sample distribution of $J(\gamma)$ is F with N and $(T - N - 1)$ degrees of freedom (see Gibbons, Ross, and Shanken [1989]). In large samples $J(\gamma)$ will have a non-central χ^2 distribution, and the assumption of normality of ϵ_τ can be dropped.

We will work with the finite-sample (non-central) χ^2 distribution, but our analysis applies to the finite sample F distribution and the large-sample χ^2 distribution, *mutatis mutandis*. The statistical properties hold for all values of γ.

1.3.2 GRS Test; $\gamma = 0$

The γ-GRS test for $\gamma = 0$ deserves special attention. It is the original GRS test developed in Gibbons, Ross, and Shanken [1989]. When $\gamma = 0$ the projection presented in (1.4) becomes a projection of assets' excess returns with respect to the risk-free return, on the excess return of the benchmark portfolio, b.

$$\begin{aligned} & R_{n\tau} - R_{F\tau} = \alpha_n(0) + \beta_{0n}(R_{b\tau} - R_{F\tau}) + \epsilon_{n\tau} \\ & \epsilon_\tau = (\epsilon_{1\tau}, \dots, \epsilon_{N\tau}), \quad E[\epsilon_\tau] = \mathbf{0}, \quad E[\epsilon_\tau(R_{b\tau} - R_{F\tau})] = \mathbf{0}, \\ & n \in \{1, \dots, N\}, \quad \tau \in \{1, \dots, T\}. \end{aligned} \tag{1.6}$$

The test of efficiency is given by

$$\begin{aligned} & J(0) = T\left[1 + \frac{\hat{\mu}_{b0}^2}{\hat{\sigma}_b^2}\right]^{-1} \hat{\alpha}'(0)\,\Sigma^{-1}\hat{\alpha}(0) \\ & H_0: \quad \alpha(0) = \mathbf{0}. \end{aligned}$$

$J(0)$ has a non-central chi-square distribution with non-centrality parameter defined in (1.5). In this special case of $\gamma = 0$, the distribution of $J(0)$ under the null hypothesis is a chi-square with N degrees of freedom (the non-centrality parameter becomes zero).

The test estimator, $J(0)$, has a well-known financial interpretation, which can be extended to $J(\gamma)$, for $\gamma \neq 0$. This interpretation, and an interpretation of the statistic's non-centrality parameter in terms of asset correlations, are the subject of the next section.

1.4 Interpretations of the γ-GRS Test

1.4.1 Sharpe Ratios

The GRS test has an insightful financial interpretation. The $J(0)$ statistic can be split into two parts. The term $T\left[1 + \frac{\hat{\mu}_{b0}^2}{\hat{\sigma}_b^2}\right]^{-1}$ takes care of scaling, accounting for the size of the sample and the exact location of the benchmark portfolio. The term $\hat{\alpha}(0)'\,\Sigma^{-1}\hat{\alpha}(0)$ summarizes the estimated value of the intercept terms ($\alpha_n(0)$'s), accounting for the correlations between different securities. Gibbons, Ross, and Shanken [1989] show that

$$\hat{\alpha}(0)'\,\Sigma^{-1}\hat{\alpha}(0) = \left(\widehat{SR}^*\right)^2 - \left(\widehat{SR}^b\right)^2, \tag{1.7}$$

where $\widehat{SR}^*$ is the sample Sharpe ratio of the tangency portfolio in market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$. $\widehat{SR}^b$ is the sample Sharpe ratio of the benchmark portfolio. $\widehat{SR}^*$ and $\widehat{SR}^b$ are defined to be

$$\begin{aligned} & \left(\widehat{SR}^b\right)^2 = \frac{\hat{\mu}_{b0}^2}{\hat{\sigma}_b^2}, \text{ and} \\ & \left(\widehat{SR}^*\right)^2 = \hat{\mu}_0'\hat{\Omega}_{+b}^{-1}\hat{\mu}_0, \end{aligned}$$

where $\hat{\mu}_0$ and $\hat{\Omega}_{+b}$ are sample estimates of $\bar{R}_{+b}$ and Ω_{+b}. Thus,

$$\hat{\mu}_0 = \frac{1}{T}\sum_{\tau=1}^{T}(R_\tau - R_{F\tau}\mathbf{1}), \text{ and}$$
$$\hat{\Omega}_{+b} = \begin{bmatrix} \hat{\sigma}_b^2 & \hat{\sigma}_b^2\hat{\beta}_0' \\ \hat{\sigma}_b^2\hat{\beta}_0 & \sigma_b^2\hat{\beta}_0\hat{\beta}_0' + \Sigma \end{bmatrix}.$$

$J(\gamma)$, for $\gamma \neq 0$ can be analogously interpreted. Let $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$, be a market with the same set of risky assets as $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$, but where the risk-free return is given by $R_\gamma = R_F - \gamma$ instead of R_F. The theoretical and sample values of the Sharpe ratio of the tangency and the benchmark portfolio can be redefined for this market. The sample values are

$$\left(\widehat{SR}^b(\gamma)\right)^2 = \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}, \text{ and}$$
$$\left(\widehat{SR}^*(\gamma)\right)^2 = \hat{\mu}_\gamma'\hat{\Omega}_{+b\gamma}^{-1}\hat{\mu}_\gamma,$$

where

$$\hat{\mu}_\gamma = \frac{1}{T}\sum_{\tau=1}^{T}(R_\tau - R_{\gamma\tau}), \text{ and}$$
$$\hat{\Omega}_{+b\gamma} = \begin{bmatrix} \hat{\sigma}_b^2 & \hat{\sigma}_b^2\hat{\beta}_\gamma' \\ \hat{\sigma}_b^2\hat{\beta}_\gamma & \sigma_b^2\hat{\beta}_\gamma\hat{\beta}_\gamma' + \Sigma \end{bmatrix}.$$

The financial interpretation of $J(\gamma)$ is given by

$$\hat{\alpha}(\gamma)'\Sigma^{-1}\hat{\alpha}(\gamma) = \left(\widehat{SR}^*(\gamma)\right)^2 - \left(\widehat{SR}^b(\gamma)\right)^2.$$

In mean-standard deviation space, Sharpe ratios are the slopes of straight lines connecting the risk-free rate and the portfolios at hand. In the case of the GRS test, the question of efficiency of the benchmark portfolio can be interpreted as the question of whether the slope of the line passing through $(0, \hat{R}_F)$ and point $(\hat{\sigma}_b, \hat{\mu}_{b0})$, can be considered statistically equal to the slope of the sample efficient frontier. When $\gamma \neq 0$, the test statistic still compares the slopes of lines passing through the risk-free rate and either the tangency or the benchmark portfolio. However, the risk-free rate is now $\hat{R}_\gamma$, the tangency portfolio is efficient in market $\{\hat{\mu}_\gamma, \hat{\Omega}_{+b,\gamma}, \hat{R}_\gamma\}$, and the question is no longer whether these two slopes are equal. Instead, the question is whether the difference between the two slopes can be deemed statistically equal to $\gamma^2(\mathbf{1} - \beta_\gamma)'\Sigma^{-1}(\mathbf{1} - \beta_\gamma)$. The solid lines in figure 1.2 are the Sharpe ratios $\widehat{SR}^*$ and $\widehat{SR}^b$. The quantity $\hat{\alpha}(0)'\Sigma^{-1}\hat{\alpha}(0)$ is the difference between the squared slopes of these two lines. The dashed lines are $\widehat{SR}^*(\gamma)$ and $\widehat{SR}^b(\gamma)$.

From the above discussion it is clear that $J(\gamma)$ is equal to $J(0)$ in market $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$, where the risk-free rate R_F is replaced with R_γ. In other words, in constructing $J(\gamma)$ we invent a fictitious market

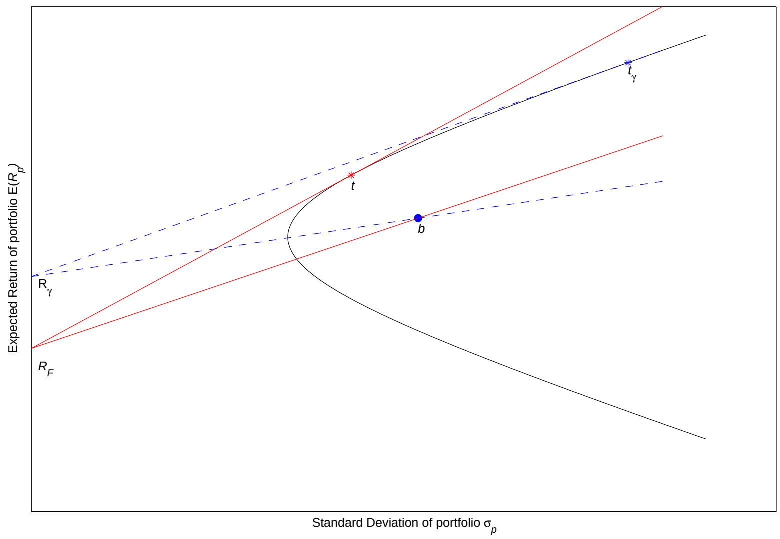


Figure 1.2. The solid lines are $\widehat{SR}^*$ and $\widehat{SR}^b$. The dashed lines are $\widehat{SR}^*(\gamma)$ and $\widehat{SR}^b(\gamma)$, for $\gamma \neq 0$.

with a fictitious risk-free rate, R_γ, and a fictitious tangency portfolio - lying on the efficient frontier for $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$ - to which the benchmark portfolio is compared. If the null hypothesis were that $\alpha(\gamma) = \mathbf{0}$, this would be a test of efficiency of portfolio b in the market $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$. That is why the null hypothesis is appropriately modified to test efficiency of portfolio b in market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$ using the tangency portfolio for market $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$.

It must also be noted that when we compare the slopes of the two dashed lines in figure 1.2, we are in fact comparing the benchmark portfolio to a new reference portfolio t_γ (the tangency portfolio in market $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$) that is not necessarily efficient in market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$. That is, the performance of the benchmark b will no longer be compared to the tangency portfolio t. If the new reference portfolio is not the best available portfolio, under the null that the benchmark portfolio is mean-variance optimal, the benchmark may sometimes outperform the (suboptimal) reference portfolio. The adjusted critical region, through adjusted null hypothesis, will reflect this.

1.4.2 Portfolio Correlations

To better understand what statistical differences there are between $J(\gamma)$ for different values of γ, it is useful to interpret the population value $\alpha'(0)\Sigma^{-1}\alpha(0)$ in terms of the correlation between the benchmark and the tangency portfolios.[4]

Let $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$ be the market from which the sample $\begin{bmatrix} R^T & R_F^T \end{bmatrix}$ is taken. We will show that

$$\alpha_0' \Sigma^{-1} \alpha_0 = (SR^*)^2 \left(1 - \rho_{tb}^2\right), \tag{1.8}$$

[4] Of course, the same interpretation holds for other values of γ. For illustration purposes it is more useful to limit attention to the interpretation of $\alpha'(0)\Sigma^{-1}\alpha(0)$; thus, the case of $\gamma = 0$.

where the absence of "hats" denotes population values, and ρ_{tb} is the correlation between the tangency portfolio t and the benchmark portfolio b, i.e.,

$$\rho_{tb} = \frac{\sigma_{bt}}{\sigma_t \sigma_b}.$$

To show (1.8), we must first note that the result in the previous subsection (see equation (1.7)) can be easily proven for population values. Thus,

$$\alpha_0' \Sigma^{-1} \alpha_0 = (SR^*)^2 - \left(SR^b\right)^2. \tag{1.9}$$

The tangency portfolio is efficient, and therefore,

$$\bar{R} - R_F \mathbf{1} = \frac{\Omega_{+b} w_t}{\sigma_t^2} (\bar{R}_t - R_F),$$

where $\frac{\Omega_{+b} w_t}{\sigma_t^2}$ is the vector of covariances of all assets in market $\{\bar{R}_{+b}, \Omega_{+b}, R_F\}$, with the tangency portfolio in that market. In particular, it is true that

$$\bar{R}_b - R_F = \frac{\sigma_{bt}}{\sigma_t^2} \left(\bar{R}_t - R_F\right). \tag{1.10}$$

Use equations (1.9) and (1.10) above, plus the definition of the Sharpe ratio of a portfolio to derive (1.8) as follows:

$$\begin{aligned}
\alpha_0' \Sigma^{-1} \alpha_0 &= (SR^*)^2 - \left(SR^b\right)^2 \\
&= (SR^*)^2 - \left[\frac{\frac{\sigma_{bt}}{\sigma_t^2}\left(\bar{R}_t - R_F\right)}{\sigma_b}\right]^2 \\
&= (SR^*)^2 - \left[\frac{\sigma_{bt}}{\sigma_t \sigma_b} \frac{\left(\bar{R}_t - R_F\right)}{\sigma_t}\right]^2 \\
&= (SR^*)^2 - \left(\frac{\sigma_{bt}}{\sigma_t \sigma_b}\right)^2 (SR^*)^2 \\
&= (SR^*)^2 \left(1 - \rho_{bt}^2\right).
\end{aligned}$$

Under the null hypothesis that $\alpha(0) = \mathbf{0}$, the distribution of the GRS test estimator is central chi-square. However, equation (1.8) tells us that when the null hypothesis is not true, the non-centrality parameter of the test statistic will depend on the covariance between the (population) return of the benchmark portfolio and the tangency portfolio. This fact means that the power of the efficiency test depends on this correlation.

Specifically, the non-centrality parameter of the distribution of $J(0)$ equals

$$\lambda = T \left[1 + \frac{\left(\hat{\mu}_{b0} - R_F\right)^2}{\hat{\sigma}_b^2}\right]^{-1} \left[(SR^*)^2 \left(1 - \rho_{tb}^2\right)\right]. \tag{1.11}$$

For a fixed critical region given by a point x_s, the area to the right under the relevant density is the power of the test. This area will normally be larger for a larger value of λ. From equation (1.11), λ *is larger for a smaller correlation between b and the tangency portfolio.*

Hence, to have a powerful test it is desirable to have a small correlation between the benchmark portfolio and the tangency portfolio. For a fixed benchmark portfolio, we will want to move the tangency portfolio to get better power. The use of a γ-GRS test is a way to take advantage of this possibility. By introducing $\gamma \neq 0$, we "move" the relevant tangency portfolio, and may improve power by doing this. The possibility of power improvement will be extensively dealt with in the next section.

1.5 Power of γ-GRS Tests

In this section we introduce new notation and definitions of power and size of the γ-GRS tests. These will be used in an example to illustrate the intuition behind the possibility of designing tests that are more powerful than the GRS test, and in our description of the power function at the end of the section.

Let $\lambda(\gamma)$ denote the non-centrality parameter of the distribution of $J(\gamma)$ when the null hypothesis is true ($\alpha_0 = \mathbf{0}$). If the null hypothesis is false, so that $\alpha(0) = c \neq \mathbf{0}$, we define

$$\tilde{\lambda}(\gamma, c) = T\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]^{-1} \left[c + (\mathbf{1} - \beta_\gamma)\gamma\right]' \Sigma^{-1} \left[c + (\mathbf{1} - \beta_\gamma)\gamma\right],$$

which is the non-centrality parameter of the distribution of $J(\gamma)$ when $\alpha(0) = c$.

Let $F(y, \lambda)$ denote the probability that a variable with a $\chi_N^2(\lambda)$ distribution (non-central chi-square distribution with N degrees of freedom and non-centrality parameter λ) be smaller than y. Let $F^{-1}(s, \lambda)$ denote the sth quantile of a $\chi_N^2(\lambda)$ distribution.

Definition 2 (Critical Region)**.** *The critical region of a statistical test is the set of values of the test statistic for which rejection of the null hypothesis is prescribed.*

In the tests we consider here, the critical region is given by a cutoff point, x_s, such that the null hypothesis is rejected if $J(\gamma) > x_s$.

Definition 3 (Size)**.** *The size of a test is the probability of rejecting the null hypothesis when it is true. Given a cutoff point x_s that defines a critical region $[x_s, \infty)$, the size of the γ-GRS test is given by*

$$1 - F(x_s, \lambda(\gamma)). \tag{1.12}$$

It is common trade to fix the size of a test and find the critical region that will deliver this size. That is, given a size, $1 - s$, a cutoff point

$$x_s(\gamma) = F^{-1}(s, \lambda(\gamma)) \tag{1.13}$$

can be found to define the critical region for the test. In equations (1.12) and (1.13), $\lambda(\gamma)$ is given by

$$\lambda(\gamma) = T\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]^{-1} \gamma^2 (\mathbf{1} - \beta_\gamma)' \Sigma^{-1} (\mathbf{1} - \beta_\gamma),$$

which is the value of the non-centrality parameter for the distribution of $J(\gamma)$ under the null hypothesis. The power of a test relates to the distribution of the test statistic under the alternative hypothesis. Since the alternative hypothesis is typically an interval, not a point, power is given by a function over this interval. We will return to this in a later subsection.

Definition 4 (Power)**.** *The power of a test is the probability of rejecting the null hypothesis when it is false. Given a critical region* $[x_s, \infty)$*, and given the true value of the Jensen* α*'s,* $\alpha(0) > 0$*, in market* $\{\bar{R}_{+b}, \Omega_{+b}, R_\gamma\}$*, the power of the* γ*-GRS test is given by*

$$1 - F\left(x_s, \tilde{\lambda}(\gamma, \alpha(\gamma))\right).$$

The following example illustrates the way in which the power and the size of the γ-GRS test relate to the value of γ.

1.5.1 Example: Power Changes with Value of γ

The example we give in this subsection shows a dramatic (unfavorable) change in the power of an efficiency test, as a product of using a specific $\gamma \neq 0$, instead of the standard $\gamma = 0$ (GRS test). In this exercise, no attempt is made to maximize power. The value of γ is chosen in the "clumsiest" possible way, to show how badly things can go if γ is poorly picked.

Consider a situation where the benchmark portfolio b is *not* efficient, but lies on the efficient frontier for the market with risky assets only, $\{\bar{R}_{+b}, \Omega_{+b}\}$. In this case, a γ^b can be found, such that the combinations of risky assets and a fictitious risk-free asset with return $R_{\gamma^b} = R_F - \gamma^b$, lie on the line tangent to the efficient frontier for market $\{\bar{R}_{+b}, \Omega_{+b}\}$ at b. Figure 1.3 illustrates this situation.

The null hypothesis ($H_0 : \alpha(0) = \mathbf{0}$) can be tested using $J(\gamma)$ for any γ. We use two tests from this family and compare their performance in terms of power. The tests we choose are $J(0)$, and $J\left(\gamma^b\right)$. We also pick the tests to have the same size $1 - s$ (in the figures we set $1 - s = 0.05$). *If the null had been true*, the distributions of $J(R_F)$ and $J\left(\gamma^b\right)$ would be given by the non-centrality parameters $\lambda(0)$ (always equal to zero), and $\lambda\left(\gamma^b\right)$ respectively.

	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = \gamma^b}$
$\boldsymbol{\alpha(\gamma)}$	$\mathbf{0}$	$\left(\mathbf{1} - \beta_{\gamma^b}\right)\gamma^b$
$\boldsymbol{\lambda(\gamma)}$	0	$T\left[1 + \frac{\hat{\mu}_{b\gamma}^2}{\hat{\sigma}_b^2}\right]^{-1} \left(\gamma^b\right)^2 (\mathbf{1} - \beta_\gamma)' \Sigma^{-1} (\mathbf{1} - \beta_\gamma)$

However, *the null is not true*. In this case, we know that $\alpha_0 = -\gamma^b\left(\mathbf{1} - \beta_{\gamma^b}\right)$ (since $\alpha\left(\gamma^b\right) = 0$ because b is an efficient portfolio in the fictitious market where R_F is replaced with R_{γ^b}, $\{\bar{R}_{+b}, \Omega_{+b}, R_{\gamma^b}\}$). As a result,

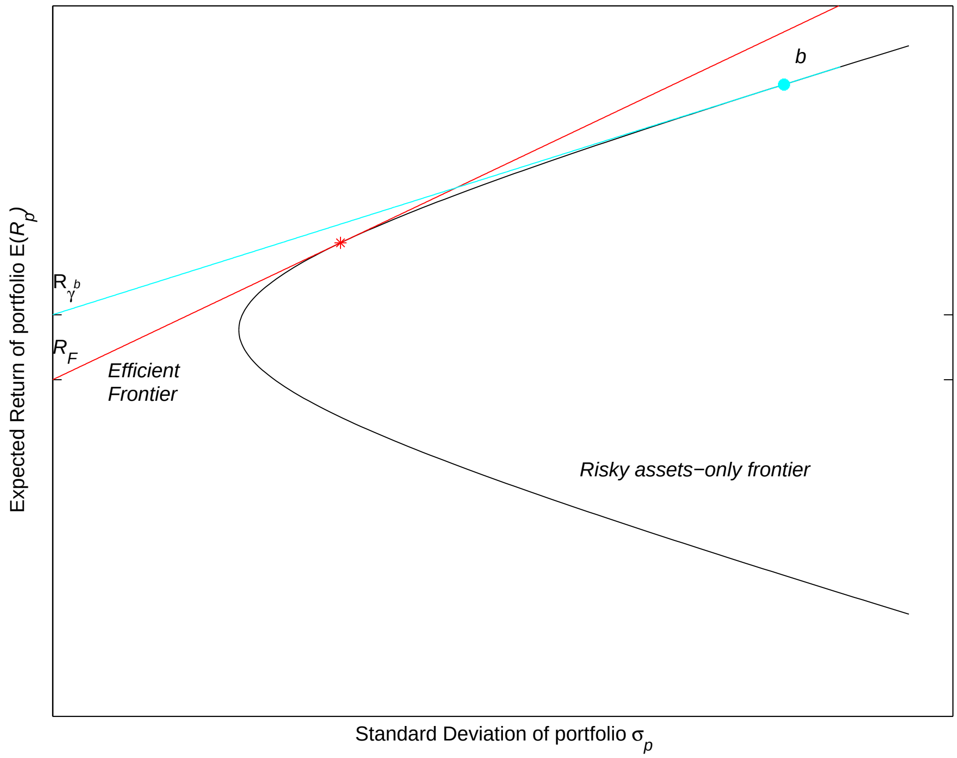


Figure 1.3. The benchmark portfolio b is not efficient, but lies on the efficient frontier for market $\{\bar{R}_{+b}, \Omega_{+b}\}$. We can draw a tangent that has intercept R_{γ^b}.

under the *true* alternative hypothesis that $\alpha_0 = -\gamma^b\left(\mathbf{1} - \beta_{\gamma^b}\right)$, the parameter $c(\gamma)$ (alternative hypothesis value of $\alpha(\gamma)$) and the non-centrality parameter are given in the following table:

	$\boldsymbol{\gamma = 0}$	$\boldsymbol{\gamma = \gamma^b}$
$\boldsymbol{c(\gamma)}$	$-\gamma^b\left(\mathbf{1} - \beta_{\gamma^b}\right)$	$\mathbf{0}$
$\boldsymbol{\tilde{\lambda}(\gamma, c)}$	$T\left[1 + \frac{\hat{\mu}_b^2}{\hat{\sigma}_b^2}\right]^{-1}\left(-\gamma^b\right)^2\left(\mathbf{1} - \beta_{\gamma^b}\right)'\Sigma^{-1}\left(\mathbf{1} - \beta_{\gamma^b}\right)$	0

The combination of the fixed size $1 - s$ and the distribution under the false null hypothesis may produce the situation depicted in figure 1.4.[5]

The *cdf* of $J(0)$ under the (false) H_0 coincides with the *cdf* of $J\left(\gamma^b\right)$ for the (true) H_1. Critical regions for both tests are determined using the distribution under the (false) null hypothesis. These regions, for size 0.05 are depicted in figure 1.4. The cutoff point for $J(0)$ lies to the left of the cutoff point for $J\left(\gamma^b\right)$, defining a probability of rejecting the null hypothesis, given that it is false, which is depicted in figure 1.5. The *cdf* of $J\left(\gamma^b\right)$ under the (false) H_0 is almost identical to that of $J(R_F)$ under H_1.[6] In figure 1.5 we can immediately see that *the test using* $J(0)$ *will be more powerful.*

We will now attempt to use the potential depicted in this example to our advantage. The objective is to

[5]We say "may" because the non-central χ^2 is not linear in the non-centrality parameter. The exact values of $\lambda(R_F)$ and $\lambda\left(\gamma^b\right)$, the degrees of freedom, and the size will jointly determine the ordering of cutoff points.

[6]To generate these plots we used CRSP data compiled by Fama and French [2004]. The specific series are monthly returns to 12 industry portfolios, returns from one-month Treasury Bills, and a portfolio of indices. The benchmark b is chosen to fit the example. Even though it seems that the *cdf* of $J\left(\gamma^b\right)$ under H_0 is identical to that of $J(0)$ under H_1, they are slightly

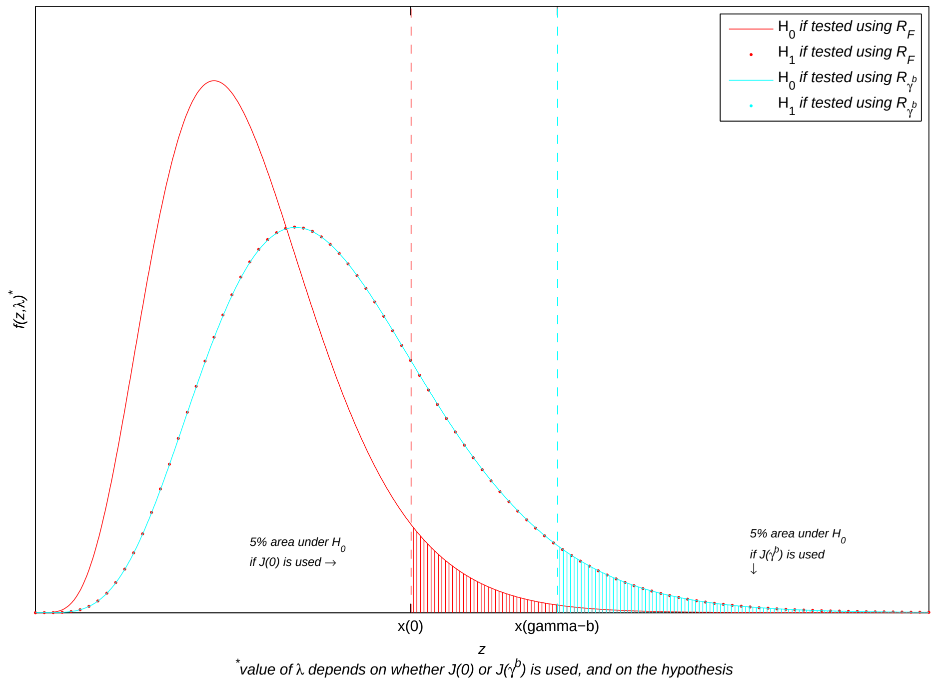


Figure 1.4. At the significance level 0.95, critical regions are displayed for $J\left(\gamma^b\right)$ and $J\left(0\right)$.

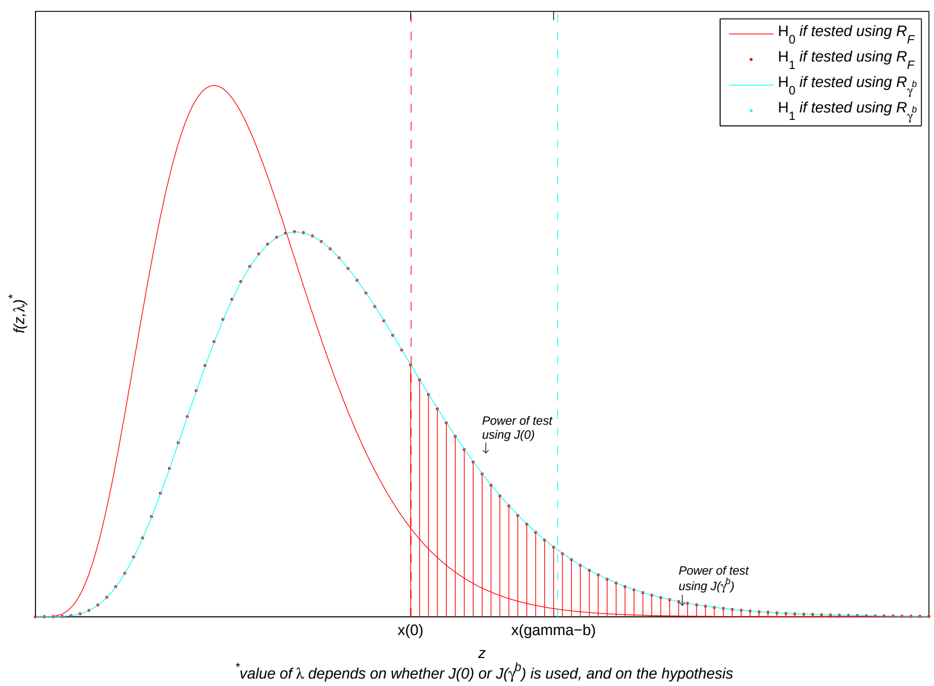


Figure 1.5. Power of the tests $J\left(\gamma^b\right)$ and $J\left(0\right)$ with size 0.05.

design a test that is more powerful than the GRS test.

1.5.2 Power Function and Power Improvement

Fix a critical region given by the cutoff point x_s. For each γ we can define the power function given the critical region to be

$$p(c; x_s, \gamma) = P(J(\gamma) > x_s | \alpha(0) = c) = 1 - F\left(x_s, \tilde{\lambda}(\gamma, c)\right), \tag{1.14}$$

where c is an arbitrary non-zero vector. That is, c is a value of $\alpha(0)$ that does not satisfy the null hypothesis. The power of a test depends on the exact *true* value of the relevant parameter ($\alpha(0)$, in this case). Since this value is unknown, power functions are defined for all possible values of the relevant parameter that correspond to the alternative hypothesis.

Equation (1.14) gives the power for a fixed cutoff point, and different values of γ and c. Notice, though, that the size of the test for a fixed cutoff point, is also changing with γ. It is more desirable to fix the size of the test, allow the cutoff point to adjust appropriately (as a function of γ), and define power as a function of size, γ, and the value c.

$$\pi(c; s, \gamma) = 1 - F\left(x_s(\gamma), \tilde{\lambda}(\gamma, c)\right),$$

where $x_s(\gamma) = F^{-1}(s, \lambda(\gamma))$. Hence, the power function is affected by the distribution both under the null and under the alternative hypotheses.

The naive approach to the problem of finding an optimal test in the γ-family of tests, is to attempt to maximize the power function stated above. This function is differentiable and locally concave, so first order conditions can be found (and second order conditions locally checked), and are stated below.

FOC:

$$\begin{aligned} 0 = & \frac{\partial F\left[F^{-1}(s, \lambda(\gamma)), \tilde{\lambda}(\gamma, c)\right]}{\partial \tilde{\lambda}(\gamma, c)} \frac{\partial \tilde{\lambda}(\gamma, c)}{\partial \gamma} \\ & + \frac{\partial F\left[F^{-1}(s, \lambda(\gamma)), \tilde{\lambda}(\gamma, c)\right]}{\partial F^{-1}(s, \lambda(\gamma))} \frac{\partial F^{-1}(s, \lambda(\gamma))}{\partial \lambda(\gamma)} \frac{\partial \lambda(\gamma)}{\partial \gamma}. \end{aligned}$$

Three points become immediately apparent from setting up the maximization problem:

1. The dependence on the alternative hypothesis, c, does not disappear at the maximum. That is, there is no γ that will maximize power for all possible points violating the null hypothesis.[7]

different. In particular: $\lambda\left(\gamma^b\right) = 4.95$, while

$$\tilde{\lambda}\left(0, -\gamma^b(\mathbf{1} - \beta)\right) = 4.93.$$

This difference is introduced by the scaling factor $T\left[1 + \frac{\left(\hat{\mu}_{b\gamma^b}\right)^2}{\hat{\sigma}_b^2}\right]^{-1}$.

[7] Notice that the notion of uniformly most powerful test in this case is different from the standard use of the term. This is so because the possibility of affecting power through γ relies on an assumption about the correlation between the benchmark

2. The first-order condition cannot be solved analytically, even for a fixed value of c. Numerical methods must be use to find maxima.

3. The fact that there is no test that maximizes power for all values of c implies that typically the GRS test will not maximize power. Moreover, the GRS need not maximize power at the *in-sample* value of portfolio correlations ($\hat{\sigma}_{bt}$).

In the absence of a γ that maximizes power for all return distributions in the alternative hypothesis, there are several options of differing validity. The approach we take in the remainder of the paper is to use a sample estimate of the true value of c and numerically determine the power maximizing γ for this value. There are several advantages of this approach, starting with its simplicity, and its asymptotic validity (plus the fact that it does not affect asymptotic distributions of the relevant statistics). Finally, for the purpose of our examples, this approach serves to illustrate that under the (implicit in all statistics) assumption that the estimated values are close to true, the GRS can have very poor power. The main disadvantages of this approach is that it uses sample information "twice," and it limits all results to one point, with no consideration of robustness. In the following section we use this approach to illustrate the potential power gains from γ-GRS tests.

1.6 Numerical Examples

The purpose of this section is to demonstrate that power improvements can indeed be achieved by using $\gamma \neq 0$. We use time series of CRSP monthly returns on 12 value-weighted industrial portfolios, taken from Fama and French [2004]. The benchmark is also taken from the Fama and French database, and is a weighted combination of the returns on NYSE, AMEX, and NASDAQ securities. We take the returns on 1-month Treasury Bills to be the risk-free rate. All series run from July 1926 to December 2003 and are given in percentage points.

We fix three levels of significance s and find the optimal test by replacing the unknown parameters with estimates. The first step of this procedure is to run the regressions specified in (1.6). Sample values of $\hat{\beta}$, $\hat{\Sigma}$, and $\hat{\alpha}\,(0)$ are taken from these regressions and inserted in the power function, which is subsequently used to find the optimal value of γ. The least-squares regression coefficients and corresponding standard deviations for each of the industrial portfolios are reported in table 1.1.

Table 1.2 reports the power maximizing γ for three levels of significance: 0.9, 0.95, and 0.99 when estimation is made over the entire sample. The average risk-free rate over the sample period is reported for comparison. The percentage increase in power of the test, when the optimal γ (denoted $\bar{\gamma}$) is used instead of the risk-free rate, is given in the last column. The reported risk-free rate is the average monthly percentage return on T-bonds. Therefore, if we think of R_γ as a fictitious risk-free rate, the reported $\bar{R}_\gamma$ will be monthly percentage returns.

portfolio and t_γ, which is an assumption about c, the true distribution of asset returns. In other words, it makes no sense to expect there to be a uniformly most powerful test in this situation.

Table 1.1. Regression (LSE) values of the parameters α_0 and β

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}
$\widehat{\beta}_n$	0.77	1.17	1.19	0.85	0.98	1.29	0.66	0.81	0.96	0.86	1.13	1.14
$\widehat{\sigma}_\beta$	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\widehat{\alpha}_{0n}$	0.19	0.12	−0.07	0.18	0.13	0.01	0.13	0.05	0.08	0.28	0.04	−0.18
$\widehat{\sigma}_{\alpha_0}$	(0.08)	(0.12)	(0.07)	(0.12)	(0.09)	(0.11)	(0.1)	(0.12)	(0.09)	(0.11)	(0.09)	(0.09)

Table 1.2. Overall maximal power values for sample 1926-2002. $\hat{R}_\gamma$ is the fictitious risk-free rate that corresponds to the power maximizing γ. The gain in power from using the power maximizing γ is also reported.

s	R_F	$\bar{R}_\gamma$	$100 \times \frac{\pi(\hat{\alpha}(0);s,\bar{\gamma})-\pi(\hat{\alpha}(0);s,0)}{\pi(\hat{\alpha}(0);s,0)}$
0.99	0.31	0.05	3.75
0.95	0.31	0.02	1.3
0.90	0.31	0.005	0.64

Figure 1.6 illustrates estimated power as a function of γ for $s = 0.99$, for the entire sample. The vertical line marks the location of R_F (since the return on T bills varies over the sample, we take R_F to be the average return on T bills) and crosses the power curve at $\pi(\hat{\alpha}_0, R_F, s)$. The maximal value of $\pi(\hat{\alpha}(0); s, \gamma)$ is achieved for $\bar{R}_\gamma = R_F - \bar{\gamma}$ smaller than R_F ($\bar{R}_\gamma = 0.05$).

Figure 1.7 shows the values of $\bar{R}_\gamma$ and R_F for 5-year periods starting June of each year between 1926 and 2002. The test has a significance $s = 0.99$, and hence its size is $1 - s = 0.01$. Figure 1.8 shows the percentage power gains achieved by using $R_{\bar{\gamma}}$ instead of R_F over the same group of 5-year periods.

The percentage power gain over the considered five-year periods peaks for the period starting in 1936, when it is 22.4%. This means that the probability of falsely accepting the mean-variance efficiency of the benchmark is reduced by as much as 18%.

1.7 An Open Question and Extensions

1.7.1 An Open Question: The Gibbons Test

The Gibbons test of efficiency is an alternative to the GRS test, which has on occasion delivered rejections of the null hypothesis when the GRS does not reject. This test assumes there is no risk-free rate, and is expected to be lenient to deliver results that do not reject the null hypothesis. It is meant for markets where there is no risk-free asset, or where this asset is unknown. When there is no risk-free rate, a result like that of equation (1.3) can be stated.

$$\bar{R}_n - \bar{R}_{zb} = \frac{\sigma_{bn}}{\sigma_b^2}\left(\bar{R}_b - \bar{R}_{zb}\right),$$

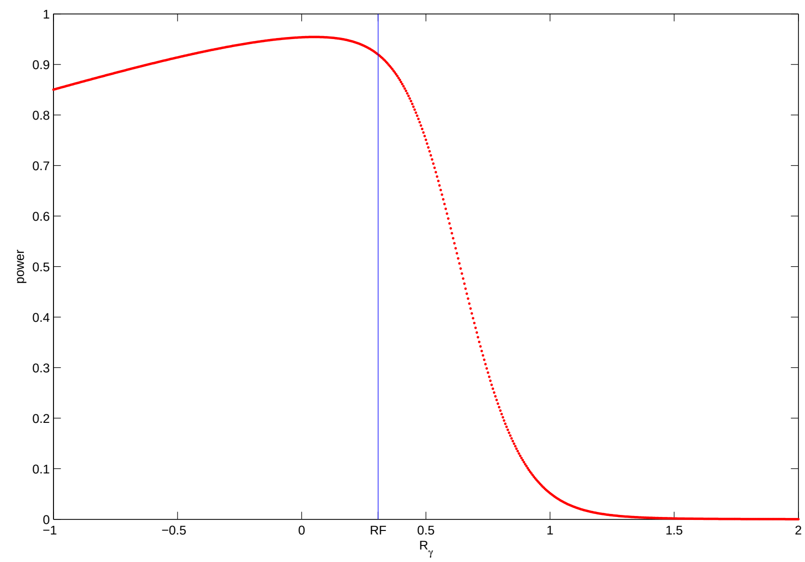


Figure 1.6. Monthly CRSP data on 12-industry portfolios, 1926-2002. Given $c = \hat{\alpha}_0$, and $s = 0.99$, we can find power as a function of R_γ.

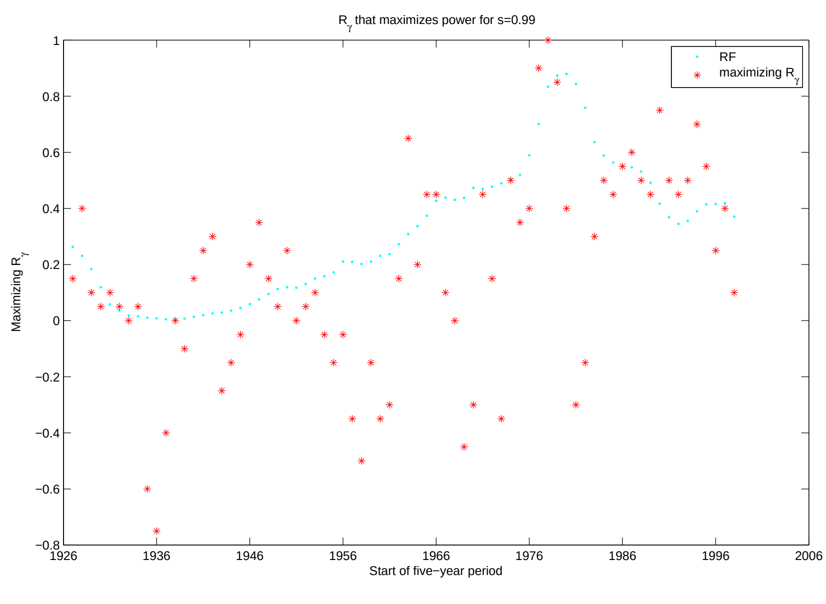


Figure 1.7. Yearly five-year periods between 1926 and 2002. The average risk-free rate and power maximizing R_γ for each period are plotted at the beginning of the period.

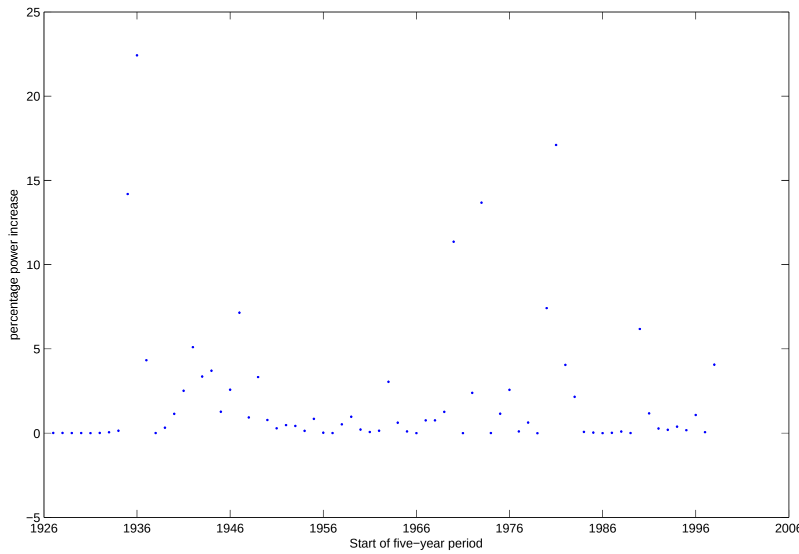


Figure 1.8. Percentage increase in power achieved using maximizing R_γ instead of R_F for yearly five-year periods between 1926 and 2002. The percentage increase corresponding to each period is plotted at the beginning of the period.

where zb denotes b's zero-covariance portfolio, which is a portfolio for which $\sigma_{b,zb} = 0$. This zero-covariance portfolio is unobservable, and $\bar{R}_{zb} = \delta$ becomes a parameter of the model. Gibbons proposes the following econometric model to estimate δ:

$$R_{n\tau} - \delta = \alpha_n(\delta) + \beta(R_{b\tau} - \delta) + \epsilon_{n\tau}$$
$$\epsilon_\tau = (\epsilon_{1\tau}, \ldots, \epsilon_{N\tau}); \quad \epsilon_\tau \sim N(\mathbf{0}, \Sigma)$$
$$n \in \{1, \ldots, N\}, \quad \tau \in \{1, \ldots, T\}.$$

The test of mean-variance efficiency of the benchmark portfolio then becomes a test of *whether there exists a* δ *such that*

$$\alpha(\delta) = (\alpha_1(\delta), \ldots, \alpha_N(\delta)) = \mathbf{0}.$$

Parameter δ can be estimated using maximum likelihood. Let δ^* denote the maximum likelihood estimate. A statistic similar to $J(\delta)$ is then computed. Denote it $J^*(\delta)$. One difference with the regression for the γ-GRS with $\gamma = \delta$ is that the fictitious risk free rate is fixed, while R_δ varies across sample points. This affects the estimated variance of the assets and portfolios. If this effect is small (for example, if the risk-free rate R_F does not vary much in time), the following discussion sheds light on the differences between the GRS and the Gibbons tests.

The main difference between the Gibbons test with intercept δ^* and $J(\delta^*)$ is that in the former the acceptance region for size $1-s$ is determined using a *central* χ^2_N distribution instead of $\chi^2_N(\lambda(\delta^*))$.[8] As a result, the Gibbons test uses the wrong critical region if indeed a risk-free rate exists. That is, its size is wrong. At the same time, since it uses a central χ^2_N distribution instead of the $\chi^2_N(\lambda(\delta^*))$, given δ^*, the power of the Gibbons test is larger than that of $J(\delta^*)$. The question remains of whether the criterion of selection of δ^* naturally leads to a correlation between the benchmark and the tangency portfolio in the market with risk-free rate δ^* that improves power. This is a relevant question we leave for future exploration.

1.7.2 Extensions

The main extension relates to the applicability of the power maximization procedure we propose and illustrate in the numerical examples. Confidence intervals for the power maximizing γ must be found and can be found through bootstrapping.

A second extension is to attempt maximization of power over subsets of the alternative hypothesis or points that are defined more abstractly than the one we use now. One such point is the null hypothesis itself. That is, one can ask the question of what γ will maximize power of the efficiency test when the truth approaches the null hypothesis. This is the *locally most powerful test* approach. The option we chose in this paper is to have a point-wise most powerful test, which is particularly useful for illustration of the power properties of the GRS test. It may be more useful in practice to have more general results regarding the direction of change in γ leading to power improvements.

[8] This is an abuse of notation. When we presented the family of tests $J(\gamma)$, γ referred to the difference between the fictitious and the true risk-free rate in the economy. Here, γ is replaced with δ^*, but δ^* is the fictitious risk-free rate, not the difference between the dummy and the real rate.

Chapter 2

Dynamically Complete Experimental Asset Markets

Abstract

We compare prices and portfolio choices in complete and incomplete experimental financial markets. The incomplete-markets treatment differs from the complete-markets one in that we shut down one market, and that we announce, halfway through trading, which of three states will not occur. The information structure in the incomplete markets is such that these markets satisfy the *necessary* condition to be *dynamically complete*. If they are indeed dynamically complete – a property that depends on the preferences of experiment participants, the individual holdings and asset prices in the incomplete-market treatment must be equivalent to those in the complete-market treatment. This is our finding. The distribution of asset holdings is undistinguishable among treatments, and state-price probability rankings coincide and are equilibrium rankings, except for one case, where incomplete markets achieve close-to-equilibrium ranking more often than complete markets.

2.1 Introduction

In this paper we investigate the experimental validity of the theory of dynamically complete asset markets. The main conclusion of this theory is that under mild assumptions, complete markets for the trade of short-lived securities, can be replicated with markets for long-lived securities, given a flow of information in time. The relevance of this conclusion comes from two sources. The first is the well-known result that complete competitive markets for aggregate-wealth risk, produce trades to efficient final outcomes. The second is the fact that the space of possible aggregate wealth outcomes is potentially very large, thus requiring an absurd number of securities to achieve the aforementioned efficiency. Dynamic completeness implies that efficiency in an uncertain world can be achieved through trade and re-trade of a small number of long-lived assets.[1]

In our experiments we compare complete and incomplete markets for the trade of the same form of aggregate risk. In the incomplete markets, securities are long-lived and information is refined in time, before the liquidation of assets. For most indicators we consider, the two setups are undistinguishable. In particular, the distribution of final individual holdings of wealth in each state of the world is undistinguishable. This is an important finding, since it is easy to manipulate participants' beliefs such that this will not be the case (see Bossaerts et al. [2007]), and "beliefs" are in fact an important part of the formation of prices in the dynamic, incomplete markets we implement experimentally.

Since the comparison of the two treatments is noisy at best, we go further, to pose the hypothesis of equilibration. If the incomplete markets are indeed dynamically complete, and equilibrium is achieved, then certain conditions must hold for prices and individual holdings, for a wide range of individual preferences. Specifically, if agents are risk averse, state-price probabilities (the prices of Arrow-Debreu securities that are implied by the prices of the tradable assets) must be ranked in inverse order of aggregate wealth. That is, the price of consumption in a *wealthy* state of the world, must be lower than the price of consumption in a *poor* state of the world. We find that state-price probabilities in the incomplete-market treatment, are ranked in accordance to the above criterion. In the complete-market treatment, state-price probabilities are mostly correctly ranked, but a systematic divergence from this hypothesis arises, which we discuss in the paper.

In both treatments we consider, there are three states of the world. In the complete-market treatment there are three securities with linearly independent distribution of dividends. There is one period for the trade of these securities, after which the securities are liquidated. In the incomplete-market treatment, although participants are endowed with identical holdings of the three securities traded in the complete market, they are allowed to trade only two of them. This is an insufficient number for the market to be complete. However, in this treatment there are two trading periods. One initial period happens with the original information about the distribution of states of the world. The second period happens after an announcement is made that correlates with the true realized state of the world, and thus, refines the information of the participants. In fact, the announcement reduces the number of possible states of the world to two.

[1] How small, and how much smaller than the number of short-lived securities, depends on the nature of the flow of information about aggregate risk in the market.

The markets *after announcement* are always complete. However, whether the market before announcement is such that the entire situation is dynamically complete, depends on the prices at which participants trade in the second period. This means that whether this treatment is dynamically complete or not, depends on the preferences of participants. A result by Kreps [1982] allows us to rule out the possibility that our incomplete-market treatment is *not* dynamically complete, as a knife-edge situation. The theoretical predictions we test here are also summarized in this work, Kreps [1982], for discrete-time markets and information, as is the case in our experimental situation. The theory is developed for continuous-time information and price processes, in Duffie and Huang [1985].

It is important to note that the precise knowledge of the information flow and the distribution of dividends, are key to the tests we perform in this paper. These are privileges of the experimental paradigm. Two important features were taken into account when designing the experiment we described in what follows. The first is that the computation of state-price probabilities requires more information than is available in any single experimental period. Aggregate uncertainty is manipulated to bypass this constraint and compute state-price probabilities with the information of one period only. The second is that the assumption of mean-variance preferences, which is used in setting up some of our hypotheses, may imply efficiency *without* completeness (or dynamic completeness). Again, the experiment is designed such that the assumption of mean-variance efficiency does not trivially imply the validity of the tests we use (the tests relate to efficiency) and thus, render the exercise meaningless.

The following section introduces the theoretical background and notation used in the remainder of the paper. It also contains the description of the experiment, both in a separate subsection, subsection 2.2.3, and in examples used to illustrate the notation. Section 2.3 presents and discusses experimental results that relate to prices and state-price probabilities. Section 2.4 presents and discusses results that relate to individual asset holdings. Section 3.5 concludes.

2.2 Notation and Experimental Setup

In this section we give the setup necessary to describe the experiment and hypotheses that are the focus of this paper. A general framework for the study of dynamic completeness in discrete-time multi-period markets can be found in Kreps [1982]. We give a more specific description, where we include probabilities for information sequences, since these are part of the experimental setup.

2.2.1 Notation

We consider asset markets for long-lived securities that liquidate in consumption period T, after T trading periods, 0 to $T-1$. Periods are indexed $t \in \{0, \ldots, T\}$. There is a single consumption good (e.g., money), which is also the numeraire for asset dividends and prices. An asset market is defined by a matrix and probability distribution over asset liquidating dividends, a vector of total asset supply, a set of possible information sequences, and a statement of what assets can be traded. We will restrict attention to dynamic markets with a simple linear information structure that we specify below. An asset market is thus defined by

$$\{\tilde{R}, R, w, \mathcal{M}, \pi_0, \}, \tag{2.1}$$

where $\tilde{R}$ is a $S \times \tilde{K}$ matrix of risky asset dividends. There are $\tilde{K}$ risky assets paying dividends that depend on the realization of a random variable called *state of the world*. There are S possible realizations of the state of the world, denoted $s \in \{1, \ldots, S\}$. The vector $w = (w_1, \ldots, w_{\tilde{K}})$ identifies supply of each asset. Let $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_S)$ be the *endowment* or total supply of the unique consumption good in each state of the world. Clearly, $\tilde{w} = \tilde{R}w$. The matrix R of dividends of tradable assets, has dimension $S \times K$, $K \leq \tilde{K}$. Without loss of generality, we assume that whenever $K < \tilde{K}$, securities $\{1, \ldots, K\}$ are tradable, while securities $\{K+1, \ldots, \tilde{K}\}$ cannot be traded.

$\mathcal{M}$ is a set of possible sequences that define how information evolves in time. Each sequence $\{M_t\}_{t=1}^T \in \mathcal{M}$, has an ex-ante probability of occurrence, $\pi_0\left(\{M_t\}_{t=1}^T\right)$. Sequences are mutually exclusive. The probability distribution over final realizations of the state of the world is implicit in the structure of $\mathcal{M}$. At every period $t > 0$ M_t determines the "surviving" return matrix, R_t, from the previous period return matrix, R_{t-1}, through the simple formula

$$R_t = M_t R_{t-1}.$$

Information at time $t = 0$ is always $R_0 = R$, and at time $t = T$, uncertainty is resolved in a single state of the world. That is, $R_T = r_s$, where r_s is the sth row of matrix R, for some $s \in \{1, \ldots, S\}$. Every matrix M_t is composed of zeros and ones. Along a sequence in $\mathcal{M}$, the number of rows and columns of M_t is non-increasing in t, and M_T always has only one row. Each column of a matrix M_t is either a column of zeros or a basis vector e_i, containing a 1 in the ith row, and zeros everywhere else. M_t contains zero-vectors in column-positions corresponding to the states "eliminated" at time 1 along the information sequence being considered. The remaining columns are $e_1, \ldots, e_{S_t}$, entered in ascending index order (skipping the positions where there is a vector of zeros).

The distribution π_0 defines the ex-ante probabilities of occurrence of each sequence in $\mathcal{M}$. Probabilities can be updated in the obvious way, as time and information evolve. We will use $\{M_\tau\}_{\tau=1}^t$ to denote the statement, "up to time t information has followed the path given by $\{M_1, \ldots, M_t\}$."

Example 1 (Experimental markets)**.** *In our experiment we consider two markets. In both cases there are three states of the world with names "X", "Y", and "Z", and three securities called A,* Bond*, and B. The markets differ in the number of periods and the tradable securities.*

Using the notation given above, the so-called complete market *treatment has $T = 1$ and is described as follows:*

1. $\tilde{R} = R = \begin{bmatrix} 0.5 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0.5 \end{bmatrix}.$

2. $w = (8, 4, 2) \Rightarrow \tilde{w} = (6, 6, 12)$.[2]

3. $\mathcal{M} = \{\begin{bmatrix}1 & 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0 & 1\end{bmatrix}\}$.

4. $\pi_0 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.[3]

The incomplete-market *treatment has $T = 2$ and is described as follows:*

1. $\tilde{R} = \begin{bmatrix}0.5 & 1 & 0\\ 0 & 1 & 1\\ 1 & 1 & 0.5\end{bmatrix}$, $R = \begin{bmatrix}0.5 & 1\\ 0 & 1\\ 1 & 1\end{bmatrix}$.

2. $w = (8, 4, 2) \Rightarrow \tilde{w} = (6, 6, 12)$.

3. $\mathcal{M} = \{\{M_{11}, M_{21}\}, \{M_{11}, M_{22}\}, \{M_{12}, M_{23}\}, \{M_{12}, M_{24}\}\}$, *where*

$$M_{11} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}, M_{12} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix},$$

and

$$M_{21} = M_{23} = \begin{bmatrix}1 & 0\end{bmatrix}, M_{22} = M_{24} = \begin{bmatrix}0 & 1\end{bmatrix}.$$

In words, the first sequence in $\mathcal{M}$ corresponds to "at $t = 1$ it is revealed that the state is not Z, and at time $t = 2$ it is revealed that the state is X". The other sequences have analogous meanings

4. $\pi_0 = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}\right)$.

Let $\mathcal{M}^s \subset \mathcal{M}$ be the set of all sequences such that $R_T = M_T R_{T-1} = r_s$. The ex-ante ($t = 0$) probability of occurrence of state of the world s is inferred from the probability over information-transformation sequences as follows:

$$\pi^s = \sum_{\{M_t\}_{t=1}^T \in \mathcal{M}^s} \pi_0 \left(\{M_t\}_{t=1}^T\right). \quad (2.2)$$[4]

For $t > 0$, a pair $(t, \{M_\tau\}_{\tau=1}^t)$ is called a *node*. At $t = 0$ there is a unique *initial node* denoted 0. The successors of node 0 are all the period-one nodes defined by sequences in $\mathcal{M}$. That is,

$$\Phi(0) = \{(1, M_1) \,|\{M_1, \ldots, M_T\} \in \mathcal{M}\}.$$

For $t > 0$, the set of *successors* of a node, $\left(t, \{\bar{M}_\tau\}_{\tau=1}^t\right)$ is defined as follows:

$$\Phi\left(t, \{\bar{M}_\tau\}_{\tau=1}^t\right) = \{\left(t+1, \{M_\tau\}_{\tau=1}^{t+1}\right) |\{M_\tau\}_{\tau=1}^t = \{\bar{M}_\tau\}_{\tau=1}^t \text{ and } \{M_\tau\}_{\tau=1}^T \in \mathcal{M}\}.$$

[2]The market portfolio differed across experimental sessions. Exact numbers are given in Table 2.3.

[3]During an experimental session, states of the world were drawn without replacement, causing variation in the probability distribution. We explain this further in the next section. Probabilities in all periods remained close to uniform across states of the world.

[4]It is evident from this statement that our notation is inconsistent with the general treatment of information sequences as sequences of finer partitions of the state space. In that treatment, each terminal node must correspond to a separate state. In our example, that would mean that there is an additional state, call it Y', in which the assets pay the same dividends as in state Y.

Example 2. *In both treatments of our experiment, the probabilities of final states of the world are* $\left(\pi^X, \pi^Y, \pi^Z\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

In the incomplete market treatment, we have the following sets of successors:

$$\Phi(0) = \{(1, M_{11}), (1, M_{12})\}$$
$$\Phi(1, M_{11}) = \{(2, \{M_{11}, M_{21}\}), (2, \{M_{11}, M_{22}\})\}$$
$$\Phi(1, M_{12}) = \{(2, \{M_{12}, M_{23}\}), (2, \{M_{12}, M_{24}\})\}.$$

The complete market treatment has only one set of successors, that of the matrix R, which is composed of the three vectors r_s.

Let $\{q(t, \{M_\tau\}_{\tau=1}^t)\}_{t=0}^{T-1}$ be a sequence of prices of assets defined for every node in market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0, \}$. $q(t, \{M_\tau\}_{\tau=1}^t)$ is a $K \times 1$ vector of prices for each tradable asset. We can now define market *completeness* and *dynamic completeness.*

We say that the asset market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ is **complete** if rank $R = S$.

We say that the asset market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with asset prices $\{q(t, \{M_\tau\}_{\tau=1}^t)\}_{t=0}^{T-1}$ is **dynamically complete** if, for every sequence $\{M_t\}_{t=1}^T$, the following statements are true:

1. rank $R_{T-1} = S_{T-1}$, where $R_{T-1} = \left(\Pi_{t=1}^{T-1} M_t\right) R$.

2. For $t \in \{0, \ldots, T-2\}$, the price vectors for every successor of $(t, \{M_\tau\}_{\tau=1}^t)$ are linearly independent. Let

$$q\left(\Phi\left(t, \{M_\tau\}_{\tau=1}^t\right)\right) = \begin{bmatrix} q(t+1, \{M_1, \ldots, M_t, M_{t+1}\}) \\ \cdots \\ q\left(t+1, \{M_1, \ldots, M_t, M'_{t+1}\}\right) \end{bmatrix},$$

be the matrix of price vectors for all successors of $(t, \{M_\tau\}_{\tau=1}^t)$. The requirement that these prices be linearly independent is equivalently stated as

$$\text{rank } q\left(\Phi\left(t, \{M_\tau\}_{\tau=1}^t\right)\right) = |\Phi\left(t, \{M_\tau\}_{\tau=1}^t\right)|.$$

An asset market, $\{\tilde{R}, R, w, \mathcal{M}, \pi_0, \}$, has an associated *Arrow-Debreu market,*

$$\{I^S, \tilde{w}, \mathcal{M}, \pi_0\},$$

where I^S denotes the $S \times S$ identity matrix, indicating that the tradable assets are units of contingent consumption for each state of the world.[5]

Consumers embedded in a market structure $\{I^S, \tilde{w}, \mathcal{M}, \pi_0\}$ will be able to achieve efficient consumption in equilibrium, in spite of the uncertainty about future supply of the consumption good. However, it may

[5] We omit the matrix of tradable assets in the description, because in the Arrow-Debreu markets all assets given in I^S can be traded.

be the case that S is very large, thus implying a very large number of tradable securities to achieve efficient outcomes. The main result of **?** is that, given the information structure implied by $\mathcal{M}$, efficient consumption in equilibrium can be achieved in a *dynamically complete* asset market with K significantly smaller than S.

Let there be N consumers indexed by $n \in \{1, \ldots, N\}$. Consumer n has preferences v_n over consumption at time T. Each consumer is endowed with $\tilde{w}_{ns} \geq 0$ units of the consumption good in state of the world s. Consumers are embedded in an Arrow-Debreu market structure with $\tilde{w} = \sum_{n=1}^{N} \tilde{w}_n$ (where $\tilde{w}_n = (\tilde{w}_{n1}, \ldots, \tilde{w}_{nS})$).

Definition 5. *An allocation* $(x_1^*, \ldots, x_N^*) \in \Re^{N \times S}$ *and a vector of prices for contingent consumption,* $(p_1^*, \ldots, p_S^*)$ *is an Arrow-Debreu equilibrium in market* $\{I^S, \tilde{w}, \mathcal{M}, \pi_0\}$ *if:*

i) For every n*,* x_n^* *is maximal for* v_n *in agent* n*'s budget set.*

ii) Contingent markets clear, that is

$$\sum_{n=1}^{N} x_n^* = \tilde{w}.$$

Definition 6. *Asset holdings* $\left(z_1^*\left(t, \{M_\tau\}_{\tau=1}^t\right), \ldots, z_N^*\left(t, \{M_\tau\}_{\tau=1}^t\right)\right)$ *and asset prices* $q^*\left(t, \{M_\tau\}_{\tau=1}^t\right)$*, constitute a Radner equilibrium in market* $\{\tilde{R}, R, w, \mathcal{M}, \pi_0, \}$ *if:*

i) For every n*,* $\tilde{R} z_n^*\left(T-1, \{M_\tau\}_{\tau=1}^{T-1}\right)$ *is maximal for* v_n *in agent* n*'s budget set, in every state of the world, and under the condition that* $z_{nk}^*\left(t, \{M_\tau\}_{\tau=1}^t\right) = w_{nk}$ *for* $k > K$ *(non-tradable assets), and all* n, t*.*

ii) At every node, asset markets clear,

$$\sum_{n=1}^{N} z_n^*\left(t, \{M_\tau\}_{\tau=1}^t\right) = w.$$

2.2.2 State-Price Probabilities

If the asset market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with prices $q^*\left(t, \{M_\tau\}_{\tau=1}^t\right)$, is dynamically complete, then the prices, p^*, and contingent consumption, x_n^*, can be uniquely determined from $q^*(\cdot)$ and $z_n^*(\cdot)$, respectively.

Moreover, if market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with prices $q\left(t, \{M_\tau\}_{\tau=1}^t\right)$ (not necessarily equilibrium prices), is dynamically complete, then a unique vector of prices of Arrow-Debreu securities can be determined from the prices $q\left(t, \{M_\tau\}_{\tau=1}^t\right)$. This vector satisfies

$$q_0\left(R, \pi_0\right) = R' p, \tag{2.3}$$

and is called the vector of *state-price probabilities*.

If, in addition, market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ is complete, p can be uniquely determined as the solution to equation (2.3), since R is invertible.

Prices of contingent consumption can be determined from asset prices at every node. We will call these *conditional* state prices, and denote them $p\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$. An equation analogous to equation (2.3) holds for conditional state prices.

We have not yet been specific about the nature of consumers' preferences, v_n. The hypotheses that direct our experimental findings are based on specific assumptions about v_n. First we assume v_n admits an expected utility representation and that all consumers are risk averse (decreasing marginal utility over final consumption). Under these assumptions, the ranking of prices in an Arrow-Debreu equilibrium is fully determined by the ranking of endowments in different states of the world.

Theoretical Prediction 1: Given an Arrow-Debreu market $\{I^S, \tilde{w}, \mathcal{M}, \pi_0\}$ and risk averse expected utility-maximizing consumers, the Arrow-Debreu equilibrium prices of consumption in every state of the world, p^* satisfy

$$\tilde{w}_s > \tilde{w}_{s'} \Rightarrow \frac{p^*_{s'}}{\pi^{s'}} > \frac{p^*_s}{\pi^s}, \tag{2.4}$$

for all $s, s' \in \{1, \ldots, S\}$.

The above also holds for Arrow-Debreu equilibrium prices determined for nodes different from 0, or *conditional state prices.* As mentioned before, if the asset market $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with equilibrium prices $q^*\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$ is dynamically complete, state-price probabilities and conditional state-price probabilities can be uniquely determined, and will satisfy the same ranking condition.

Theoretical Prediction 2: If $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with equilibrium prices $q^*\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$ is dynamically complete and consumers are risk-averse, expected utility maximizers, the associated state-price probabilities and conditional state-price probabilities, $p^*\left(t, \{M_\tau\}_{\tau=1}^{t}\right), t \in \{0, \ldots, T-1\}$ satisfy

$$\tilde{w}_s > \tilde{w}_{s'} \Rightarrow \frac{p^*_{s'}\left(t, \{M_\tau\}_{\tau=1}^{t}\right)}{\pi^{s'}\left(t, \{M_\tau\}_{\tau=1}^{t}\right)} > \frac{p^*_s\left(T, \{M_\tau\}_{\tau=1}^{t}\right)}{\pi^s\left(t, \{M_\tau\}_{\tau=1}^{t}\right)}, \tag{2.5}$$

for every $s, s' \in \{1, \ldots, S\}$ such that $\pi^s(\cdot) > 0, \pi^{s'}(\cdot) > 0$, and where $\pi^s\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$ denotes the updated probability of state s at node $\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$.

Analogous results hold for asset holdings. We present them as theoretical predictions 3 and 4 below.

Theoretical Prediction 3: If $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with prices $q^*\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$ is dynamically complete, then the Radner equilibrium holdings, $z^*_n\left(t, \{M_\tau\}_{\tau=1}^{t}\right)$ and the Arrow-Debreu equilibrium holdings of the associated Arrow-Debreu market, satisfy:

$$\tilde{R} z^*_n\left(T-1, \{M_\tau\}_{\tau=1}^{T-1}\right) = x^*_n, \forall n. \tag{2.6}$$

The next prediction requires a much tighter restriction of consumers' preferences.[6] Under the assumption that consumers' preference-maximizing choice of asset holdings equals the choice of an agent maximizing mean-variance utility over asset holdings, in equilibrium all consumers will have relative holdings of assets

[6]It is not strictly correct to call this a restriction on preferences, since this restriction on representation of consumers' choices of asset holdings depends both on preferences and the probability distribution of states of the world (see Berk [1997] for a thorough discussion of the subject).

equal to the market portfolio (if all assets can be traded). The *market portfolio* in the Arrow-Debreu market, is given by $\tilde{w}$, while it is w in the asset market.

Theoretical Prediction 4: If consumers have preferences that can be represented with mean-variance utility over asset holdings, then the Arrow-Debreu equilibrium holdings satisfy

$$\frac{x^*_{ns}}{x^*_{ns'}} = \frac{\tilde{w}_s}{\tilde{w}_{s'}}, \forall s, s' \in \{1, \ldots, S\}.$$

If $\{\tilde{R}, R, w, \mathcal{M}, \pi_0\}$ with prices $\{q^* \left(t, \{M_\tau\}_{\tau=1}^t\right)$ is dynamically complete, then the Radner equilibrium holdings, $z^*_n \left(t, \{M_\tau\}_{\tau=1}^t\right)$, satisfy:

$$\frac{\sum_{k=1}^{\tilde{K}} r_{sk} z^*_{nk} \left(T-1, \{M_\tau\}_{\tau=1}^{T-1}\right)}{\sum_{k=1}^{\tilde{K}} r_{s'k} z^*_{nk} \left(T-1, \{M_\tau\}_{\tau=1}^{T-1}\right)} = \frac{\tilde{w}_s}{\tilde{w}_{s'}}, \forall s, s' \in \{1, \ldots, S\}, \tag{2.7}$$

where r_{sk} is the dividend on asset k in state of the world s.

Example 3 (State-price probabilities ranking in experiment)**.** *The markets considered in the experiment have total supply of consumption good, $\tilde{w} = (6, 6, 12)$. Given that all three states have equal probability of occurrence , the Arrow-Debreu equilibrium prices should be ranked as follows:*

$$p^*_X = p^*_Y > p^*_Z,$$

where we have replaced the numerical indices with letters denoting the states. Moreover,

$$\frac{3}{2} p^*_X \left(1, \{M_{11}\}\right) = 3 p^*_Y \left(1, \{M_{11}\}\right), \ \textit{and} \quad 3 p^*_Y \left(1, \{M_{12}\}\right) > \frac{3}{2} p^*_Z \left(1, \{M_{12}\}\right).$$

The incomplete market treatment clearly does not implement a complete market, but prices may be such that the market is dynamically complete. If it is, then the above rankings of state-price probabilities must hold also in the incomplete market treatment.

2.2.3 Experimental Implementation of Asset Markets

In our experiments there are three securities, Asset A, Asset B, and Bonds, as well as cash. In the *complete market* treatment, all three securities can be traded by participants, while in the *incomplete market* treatment, only Asset A and Bonds can be traded. Securities can be sold short, while cash can be held only in non-negative amounts. A bankruptcy rule explained below is used to prevent agents from committing to trades that require negative cash holdings.

An experimental session lasts approximately $2\frac{1}{2}$ hours, of which approximately 30 minutes are dedicated to a review of instructions, another 30 minutes are composed of three practice periods to familiarize participants with the trading software, and the remaining time is split over 8 payoff-relevant periods. The *complete market* treatment, which serves as control, is run in three of the payoff-relevant periods, lasting 6 minutes each. The

		State of the world		
		X	Y	Z
	A	0.5	0	1
ASSET	B	0	1	0.5
	Bond	1	1	1

Table 2.1. Dollar dividends of assets, in each state of the world.

If the state is:	X	Y	Z
Then the announcement is	"The state is not Z" with probability 1	"The state is not X" or "The state is not Z" with equal probabilities	"The state is not X" with probability 1

Table 2.2. Distribution of the announcement conditional on the state of the world.

remaining periods last 8 minutes each, and are *incomplete market*-treatment periods.

At the beginning of a period, participants are endowed with holdings of the risky assets A and B, the risk-free Bonds, and cash. During a *complete market* period, participants can trade these holdings in parallel markets for all three assets, to achieve new positions. During an *incomplete market* period, participants can only trade asset A and Bonds, in two parallel markets. In both treatments trading is done through an electronic open book continuous market, or double auction, which is implemented with jMarkets software. At the end of the period, dividends of the securities are realized, and participants' earnings are added to their cumulative experimental earnings.

Dividends of assets A and B are governed by the realization of the state of the world, which can be X, Y, or Z. During an experimental session, states of the world for each period are drawn without replacement from an urn that starts the session containing 6 balls marked X, 6 marked Y, and 6 marked Z. All accounting is done in dollar units, including the values of dividends. Dividends in different states of the world were given in Example 1, and are repeated in Table 2.1 below.

At the beginning of each period, a participant knows his own holdings, the probabilities of the states of the world, and the payoff table (Table 2.1). At the beginning of an *incomplete market* period participants also know that half-way through the period (after exactly 4 minutes) they will be given a piece of information - which we will call the "announcement" - that correlates with the state of the world. The distribution of the announcement conditional on the state of the world is also known (see Table 2.2), and the distribution of the state of the world conditional on each possible announcement can be readily computed with the information available to participants. Midway an *incomplete market* period, participants learn the realization of the announcement. At the end of every period, participants learn the realization of the state of the world, their period earnings, and their total earnings for the experimental session up to that point.

Figure 2.1 illustrates the incomplete market treatment, and gives conditional probabilities for states at every node. Short-hand notation is used to denote the nodes for the experiment. Nodes $(1, M_{11})$ and $(1, M_{12})$

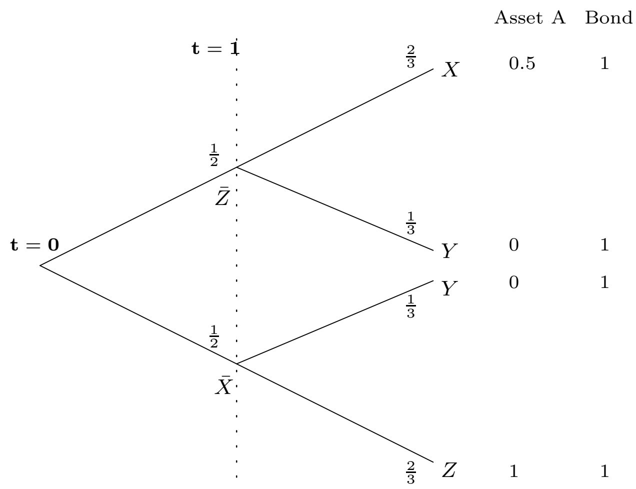


Figure 2.1. Time-event tree for the incomplete market periods.

are called $\bar{Z}$ and $\bar{X}$ respectively, referring to the announcement of what state *did not occur*.

The restriction that participants have non-negative cash holdings is implemented through a bankruptcy rule, automated and incorporated in the experimental software. Whenever a participant attempts to submit an order to buy or sell an asset, her cash holdings after dividends are computed for all states of the world, given her current asset holdings, and her outstanding orders (orders awaiting trade) that are *likely* to trade, including the order she is trying to submit. If these hypothetical cash holdings turn out negative for some state of the world, she is not allowed to submit the order.

More detailed information about the experimental setup and the software used to implement the markets, can be found in the experiment instructions on the web, at http://clef.caltech.edu/exp/dc/index.html

Two features of the experimental design deserve further comment. The first, is that we create incomplete markets by shutting down one market, not by having only two securities to start with. The second is that the expected payoff of the two risky assets in our experimental markets are equal.

Shutting down a market instead of starting with only two assets has two relevant implications. On one hand, it permits us to have identical market portfolios (not only equal state-holdings) in complete and incomplete market periods. On the other hand, by having participants hold a risky asset that they cannot trade, we can produce a situation where they cannot *trade to the market* before the announcement. They must trade after the announcement is made in order to reproduce the payoffs to the market portfolio in every state. We discuss this further in section **??**.

The second feature, the fact that assets A and B have the same expected payoff, does not mean that they should be equally priced. Asset A is twice as abundant as asset B, and has a lower price in the presence of risk aversion. This gives a clean test for the presence of risk aversion.

We report results taken from four experimental sessions, run in the Fall of 2005. Experimental sessions are identified by their date (yymmdd). Experiment 051005 had 12 participants, experiment 051022 had 22 participants, and experiments 051122 and 051201 had 31 participants each. Earnings per session per participant averaged approximately \$45, with a standard error of \$9. We give general statistics on trading behavior and prices in the Appendix.

2.3 Result: State-price probability rankings

In this section we look at theoretical predictions 1 and 2 (see Subsection 2.2.2). We first give the methodology used to compute state-price probabilities for the incomplete-market periods, given the limited information available. The methodology for computing conditional state-price probabilities in the incomplete-market periods and state-price probabilities in the complete-market periods is also briefly reviewed.

2.3.1 A Full Rank System to Compute State-Price Probabilities

If the experimental markets are dynamically complete, we know that the following holds:

$$q_0 = R'p. \tag{2.8}$$

Dynamic completeness and state-price probabilities are theoretical notions. Whether dynamic completeness is true and what state price probabilities are implied depend on all interim prices – but interim prices are not observed in an given period. One possibility to verify dynamic completeness and construct p from the sequence of asset prices is to combine asset prices from different experimental periods, where different information sequences are realized. The alternative, which we present here, is to treat the quest of computing state-price probabilities from the experimental prices as a simple algebra problem.

If the market is dynamically complete, we know that equation (2.8) holds, but state-price probabilities cannot be computed from this equation because it is an underdetermined system. In other words, since there are three possible states of the world and two independent tradeable assets, R is not invertible, and p is not uniquely defined by the above equation. However, it will be uniquely defined in the presence of one more, independent conditions to define p. Our parameter specification is such that we have an additional condition on p. Specifically, states X and Y have equal aggregate wealth. Therefore, in equilibrium, states X and Y must have equal *adjusted* state-price probabilities (the adjusted state-price probability of state s is the state-price probability of state s divided by the exogenous probability of occurrence of state s).

In other words, take a third equation:

$$\frac{p_X}{\pi^X} - \frac{p_Y}{\pi^Y} = 0 \Leftrightarrow$$

$$\begin{bmatrix} \frac{1}{\pi^X} & -\frac{1}{\pi^Y} & 0 \end{bmatrix} \begin{bmatrix} p_X \\ p_Y \\ p_Z \end{bmatrix} = 0,$$

We use the above equation to construct a new system of equations,

$$\begin{bmatrix} q_1(0) \\ q_2(0) \\ 0 \end{bmatrix} = \boldsymbol{R}p,$$

where

$$\boldsymbol{R} = \begin{bmatrix} R'_{(2\times3)} & & \\ \frac{1}{\pi^X} & -\frac{1}{\pi^Y} & 0 \end{bmatrix},$$

$q_1(0)$ is the price of asset A at time 0, and $q_2(0)$ is the price of the Bond at time 0.

The matrix $\boldsymbol{R}$ of coefficients is invertible, and p can be computed, *given that* $\frac{p_X}{\pi^X} = \frac{p_Y}{\pi^Y}$. The results we present below are computed using this method.

2.3.2 Other methodological remarks

The computation of state-price probabilities from asset prices in experimental periods with a complete market is obvious. We use equation (2.8), where R is now that corresponding to complete market periods, and is thus invertible. Solving for p is straightforward.

We proceed similarly with the computation of conditional state-price probabilities in the two conditioning nodes $(\bar{X}, \bar{Z})$. In this case, the following equation is true:

$$q(1\{M_{11}\}) = R'_{11}p$$

at node $\bar{Z}$, and an analogous equation is true at $\bar{X}$. R_{11} and R_{12} are both invertible matrices, implying that $p(1,\cdot)$ can be uniquely solved for.

In the experimental markets, trade is not instantaneous, but happens over a period of time during which the market is active. The time during which it is active before announcement is denoted $t = 0$, while $t = 1$ refers to the time when the market is open after the announcement. For each of these theoretical times there are, in the experiment, many trading prices. In describing the theoretical framework and the experiment (including the methodology), we have referred to asset prices as a single magnitude at every time. However, there are many prices at every time. In computing the state-price probabilities that we report, we will take an asset's price at a given time to be either the final price in the experimental period or the average over all trading prices in the period. We are now ready to state the first results.

2.3.3 State-Price Probabilities for All States

Figures 2.2 and 2.3 show the adjusted state-price probabilities for each period, computed using average prices - before and after announcement, for the incomplete market periods. An analogous set of figures, where adjusted state-price probabilities are derived from end-of-period prices, is presented in figures 2.4 and 2.5.

State-price probabilities are *adjusted*, meaning that they are divided by each state's probability of occurrence. This is done because the states of the world are not always equally likely, since during one experimental session, states are drawn without replacement.

Given the total supply of assets A and B, and Bonds, and given the payoff matrix of the experiment, the theoretical prediction is that states X and Y have the same adjusted state-price probability, and that this value be higher than the adjusted state-price probability of state Z.[7]

In the figures, state Z is ranked lowest in all but two experimental periods. Thus, $p_Z < p_X$ and $p_Z < p_Y$, as predicted by theory. The relative ranking of states X and Y is assumed away in periods with incomplete markets. In complete market periods the prediction that $p_X = p_Y$ does not hold. Typically, $p_Y > p_X$.

2.3.4 Ranking of Conditional State-Price Probabilities

It is perfectly possible that correct ranking of state-price probabilities is obtained in spite of incorrect ranking of conditional state-price probabilities. This can happen by coincidence - prices at $t = 0$ and at $t = 1$ violate equilibrium theory predictions, but the algebra works in such a way as to produce correct state-price probability ranking - or because participants correctly anticipate the incorrectly ranked conditional state-price probabilities, but equilibrium prices at $t = 0$ are still such that the state-price probabilities are correctly ranked.

Figures 2.6 to 2.9 show conditional state-price probabilities conditional on the announcement. They are computed using average prices after announcement. If anything, conditional state-price probability ranking is closer to the theoretical prediction than state-price probability ranking in complete market periods. In periods where the announcement is $\bar{Z}$, $\frac{p_X(\bar{Z})}{\pi^X(\bar{Z})}$ is very close to $\frac{p_Y(\bar{Z})}{\pi^Y(\bar{Z})}$, as theoretically predicted. In periods where the announcement is $\bar{X}$, $\frac{p_Y(\bar{X})}{\pi^Y(\bar{X})}$ is bigger than $\frac{p_Z(\bar{X})}{\pi^Z(\bar{X})}$, as predicted by theory.

2.4 Individual Asset Holdings

This section deals with theoretical predictions 3 and 4. In our experiments individuals do not hold the market portfolio, regardless of whether we look at a complete or an incomplete market period. However, this is not different from previous (complete market) asset pricing experiments. Moreover, Bossaerts, Plott, and

[7] This is only approximately the case in experimental sessions 051122 and 051201, where the wealth in state X is 6, it is5.8 in state Y, and 12.05 in state Z.

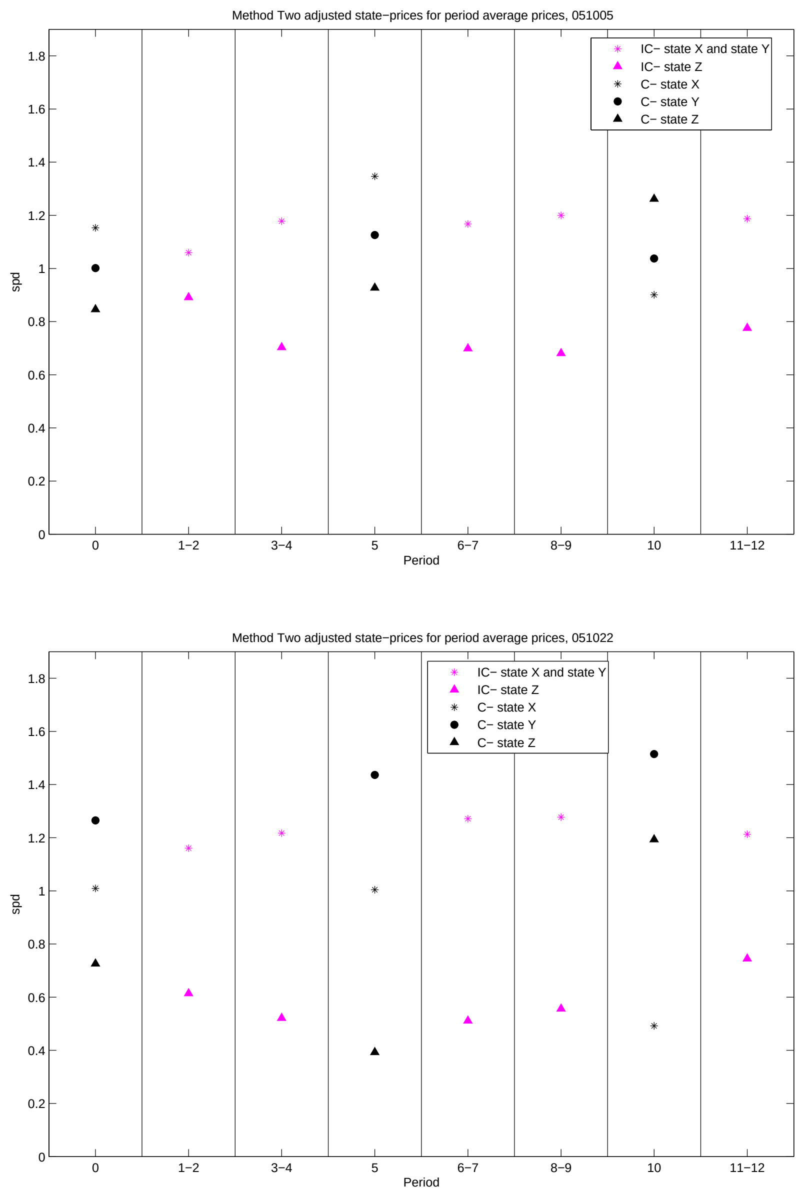


Figure 2.2. Adjusted state-price probabilities computed using average prices in each experimental period.

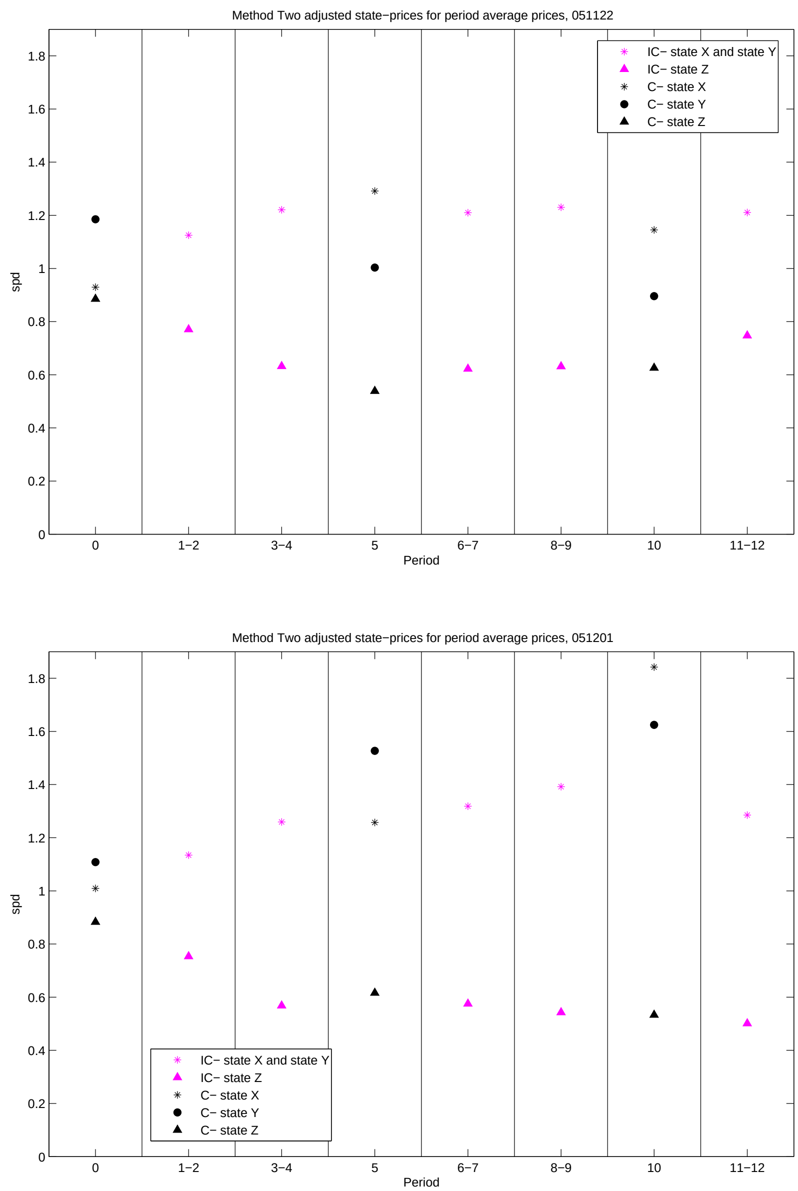
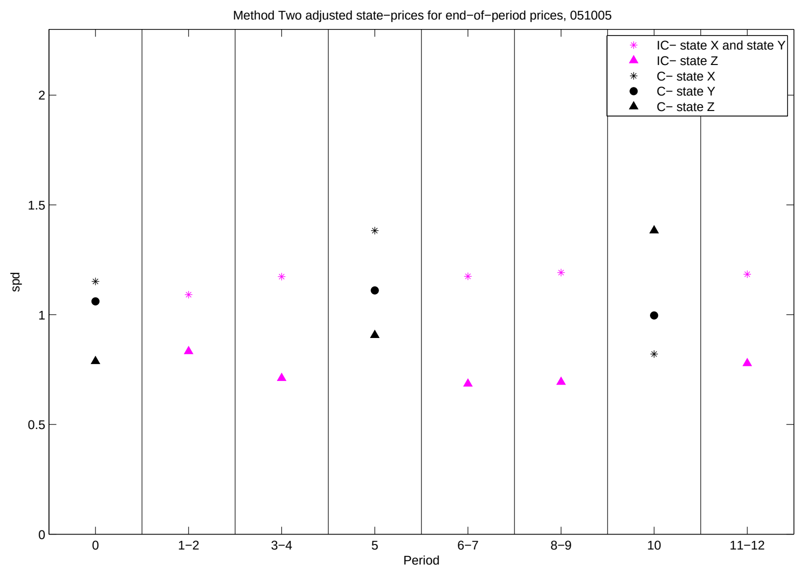


Figure 2.3. Adjusted state-price probabilities computed using average prices in each experimental period.



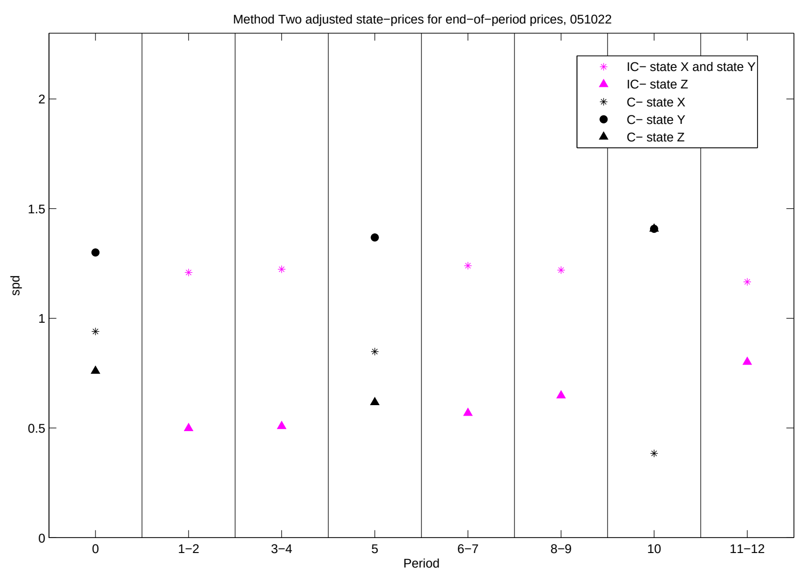


Figure 2.4. Adjusted state-price probabilities computed using final prices for each experimental period.

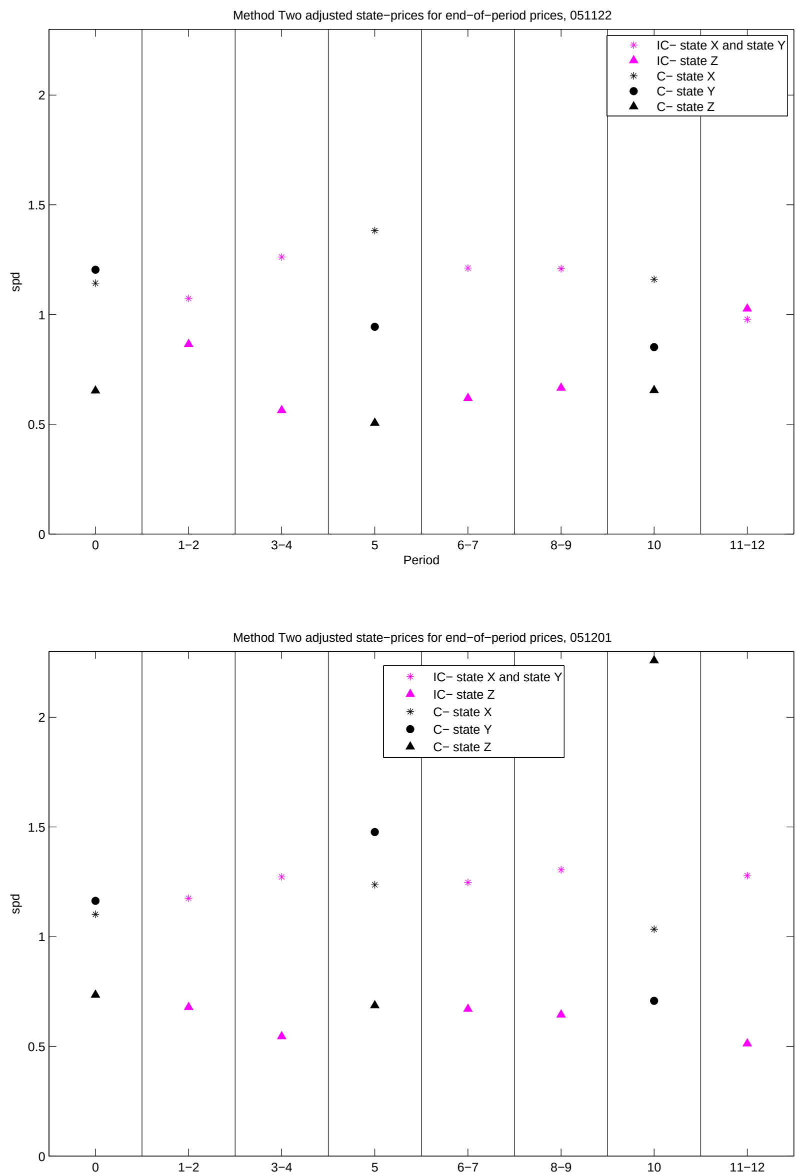


Figure 2.5. Adjusted state-price probabilities computed using final prices for each experimental period.

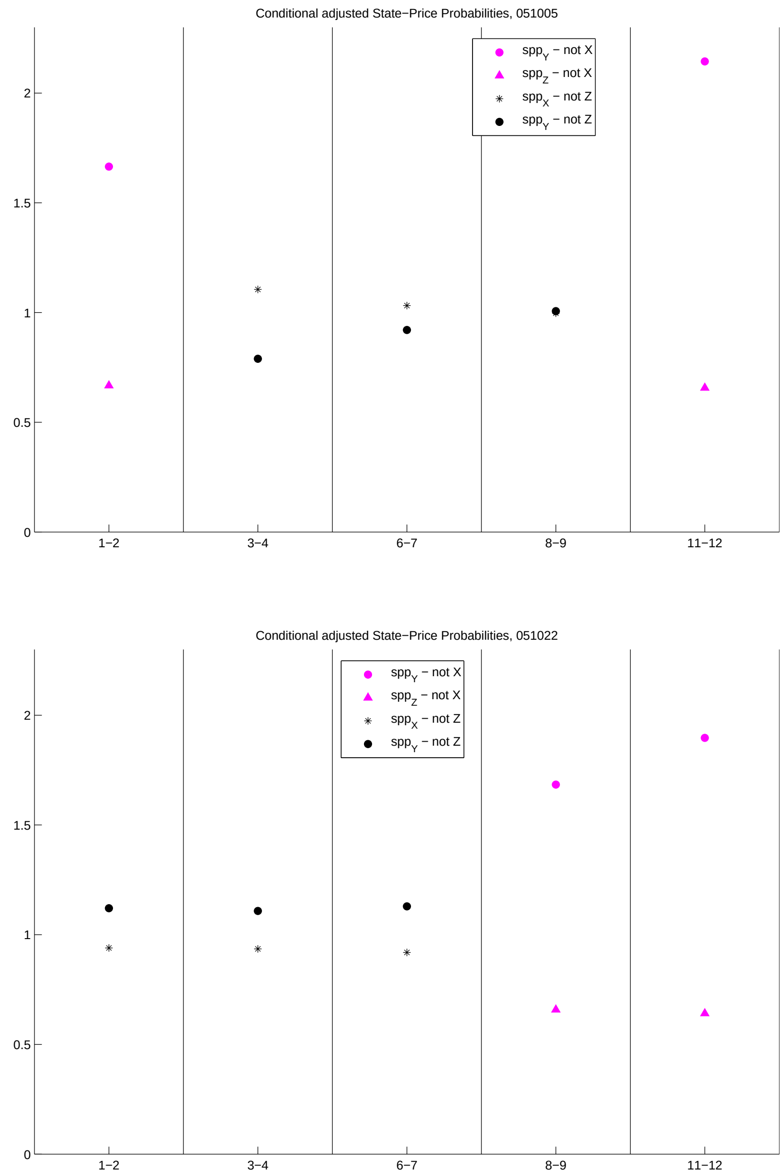


Figure 2.6. State-price probabilities conditional on announcement, computed using average prices after announcement.

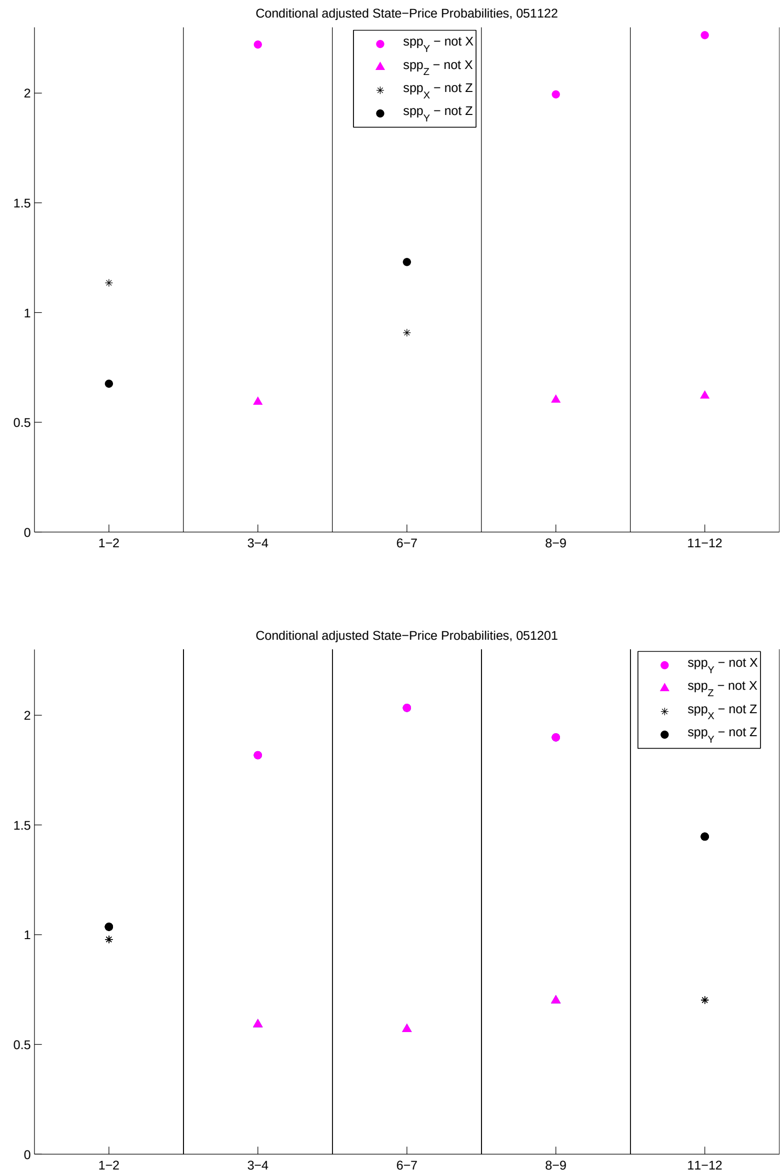


Figure 2.7. State-price probabilities conditional on announcement, computed using average prices after announcement.

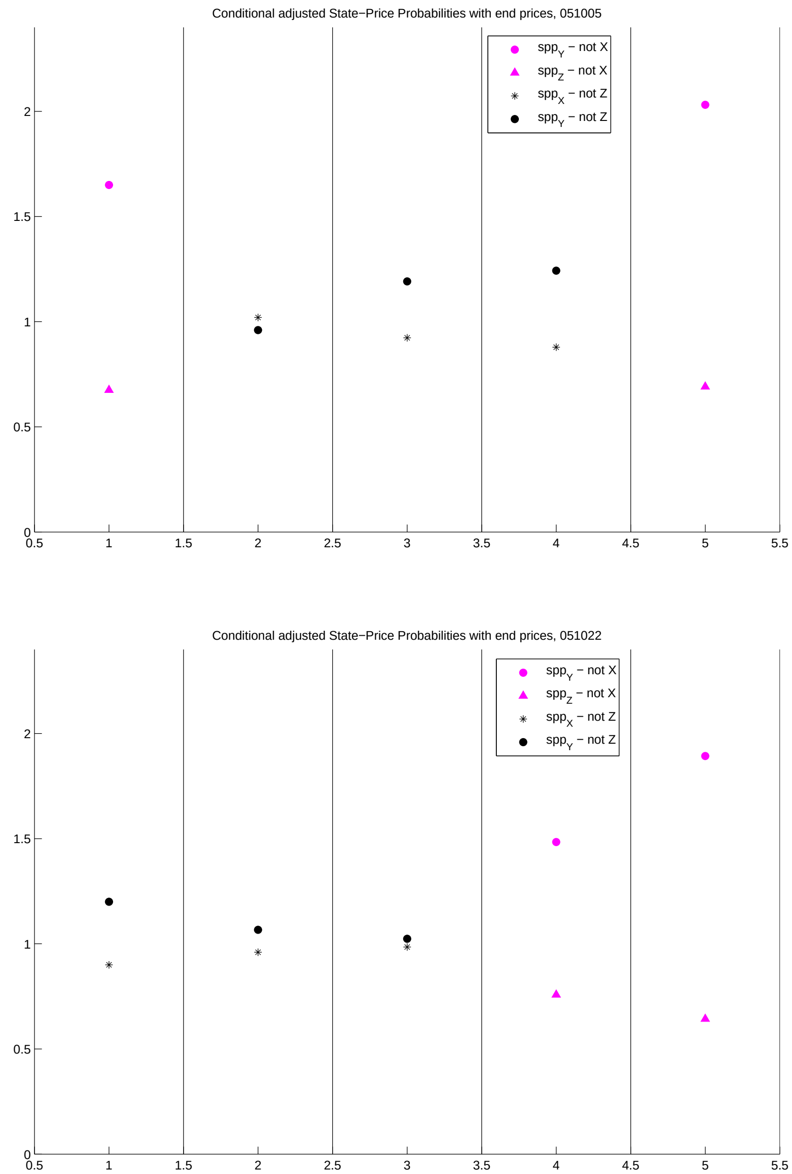


Figure 2.8. State-price probabilities conditional on announcement, computed using final prices of the period after announcement.

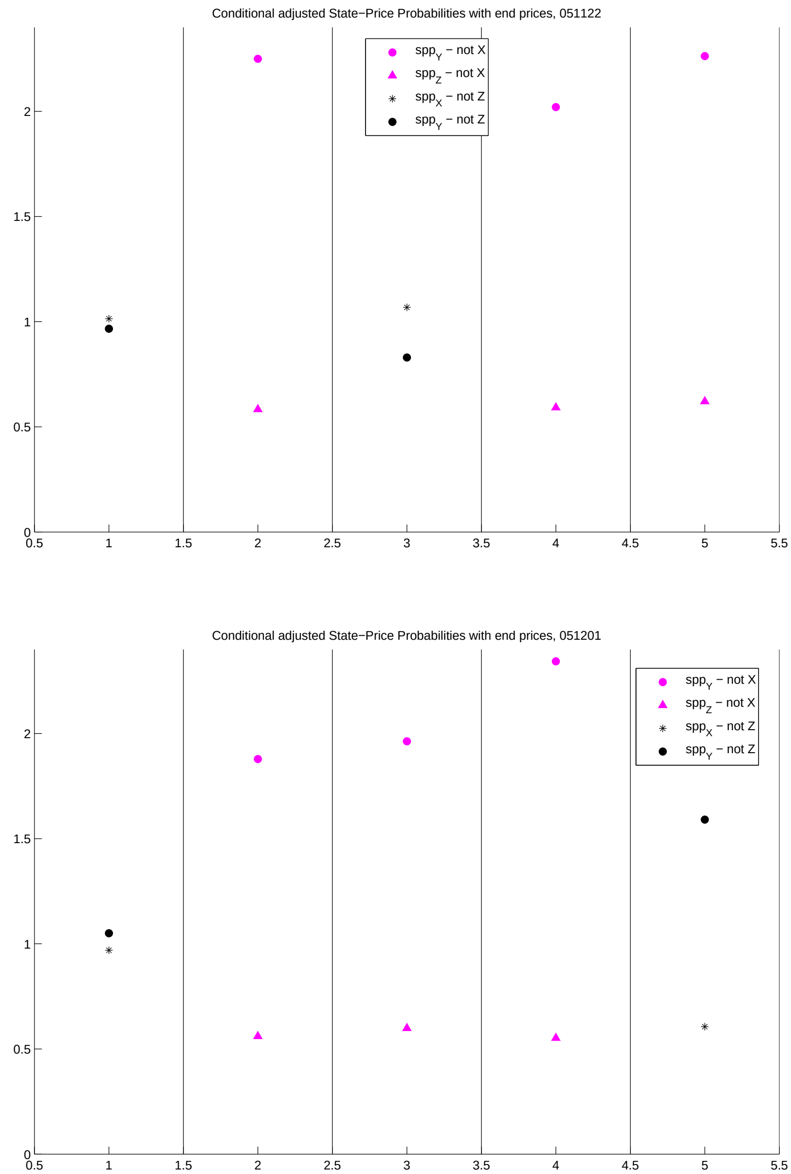


Figure 2.9. State-price probabilities conditional on announcement, computed using final prices of the period after announcement.

Initial Endowments		
	Agent type	
	I	*II*
Asset A	14	2
Asset B	0	8
Bond	0	0
Cash	$1	$3

Market Portfolio in each session			
	# agents type		
session	*I*	*II*	*Mkt. Portfolio*
051005	6	6	8A 4B 0Bonds $2
051022	11	11	8A 4B 0Bonds $2
051122	16	15	8.2A 3.9B 0Bonds $1.9
051201	16	15	8.2A 3.9B 0Bonds $1.9

Table 2.3. Market portfolio in each experimental session. This is given by the initial endowments of each agent type, and the number of subject of each type.

		State of the world		
		X	Y	Z
Type I	state-holdings	8	1	15
	rel. state-holdings	*0.333*	*0.042*	*0.625*
Type II	state-holdings	4	11	9
	rel. state-holdings	*0.167*	*0.458*	*0.375*
Market	state-holdings	6	6	12
	rel. state-holdings	*0.25*	*0.25*	*0.5*

Table 2.4. State holdings implied by the initial endowments of type I and type II participants, as well as state holdings implied by the market portfolio that obtains when there are equal numbers of participants of each type. Relative state holdings are the fraction of state holdings in one state over the sum of holdings in all states.

Zame [2005] shows that approximate CAPM equilibrium pricing (and even more so just correct state-price probability ranking) can arise in a model where individuals do not hold the market portfolio, but the *average* individual does.[8]

Figures 2.10 to 2.13 show the relative state-holdings of participants in our experiment. We call state holding the value of a participant's holdings in a given state (this is $\tilde{w}_n$ in our original notation). When $\bar{Z}$ is announced, we compute the relative state holdings for state X, $\frac{\tilde{w}_{nX}}{\tilde{w}_{nX}+\tilde{w}_{nY}}$. When $\bar{X}$ is announced, we compute the relative state holdings of state Z, $\frac{\tilde{w}_{nZ}}{\tilde{w}_{nY}+\tilde{w}_{nZ}}$. *Time-zero state holdings* can be computed with a denominator that sums over all three states. We do not report these, since they make no sense, as shown in example 4, but we will use the term in examples and tables that follow.

Before describing our findings, it is important that we illustrate what it means to trade to the market's state holdings in our setup. Table 2.3 reports the two types of initial endowments participants can hold in our experiments, and the market portfolio that follows, given the number of participants of each type we have in each experimental session. Table 2.4 shows the returns to endowments in each state (the state-holdings) for both types of initial endowments and the market, when there are equal numbers of participants of each type.

The following examples use the information in tables 2.3 and 2.4 to illustrate the exercise of "trading to the market" in our experiment.

[8]Their result says that a sample market taken from a population where the deviation of individual demands from CAPM demands has mean zero, will converge to CAPM pricing as the size of the market increases.

Example 4 (Trading to the market before announcement)**.** *This example shows that a participant of type I* cannot trade to the market before announcement*. Suppose a participant of type I tried to achieve the same state holdings as the market before the announcement, and given that she cannot trade asset B (which is the case in the incomplete market periods). The market has equal state-holdings in state X and state Y. Type I can only achieve this if she trades to* 0 *units of asset A. To see this, notice that in state Y, asset A pays* 0*, and type I has* 0 *units of asset B, which she cannot change. Hence, in state Y* all her earnings come from risk-free sources *(cash and bonds). Thus, in order to have the same holdings in state X, her earnings in state X must also come from risk-free sources only. This can only be achieved if she holds nothing of asset A. This means that type I must hold only bonds and cash (since she starts with nothing of asset B), which implies equal earnings in* all three states*. This yields the wrong relative state holdings for state Z.*

Example 5 (Trade to the market when "not Z" is announced)**.** *The market gives equal state-holdings for states X and Y. From example 4 we know that this means that type I must hold* 0 *units of asset A, no matter how much she holds of bonds and cash. She starts with* 14 *units of asset A, which she must sell to achieve the same relative state-holdings as the market.*

To have the same relative state-holdings as the market when the announcement is "not Z", agent type II must hold 16 *units of asset A (double the amount he holds of asset B, and* 8 *times his initial endowment of A!). His holdings of bonds and cash are irrelevant.*

When the announcement is "not Z", in order to equal the market's relative state holdings, both types must fine-tune their holdings of asset A with respect to their holdings of asset B, which are fixed. The following example shows that when the announcement is "not X", the important relation is that of their asset A holdings with respect to their risk-free asset holdings (cash plus bonds).

Example 6 (Trade to the market when "not X" is announced)**.** *The market holds twice as much in state Z as in state Y. Type I holds* 0 *of asset B, so all that matters are the returns to asset A in states Y and Z. Asset A yields* 0 *in state Y, and* 1 *– the same as the risk-free assets – in state Z. Knowing this, it is straightforward that she must hold the same number of units of asset A as of risk-free assets. Thus, depending on the price of asset A, type I must sell enough units of asset A to equal her holdings of cash and of asset A.*

Type II holds 8 *units of asset B, that give* \$8 *in state Y, and* \$4 *in state Z, which he cannot affect. Knowing this, it is easy to verify that his number of units of asset A must equal his number of units of risk-free assets plus* 12*. Depending on the price of asset A, he must buy asset A until this relation between bond and cash holdings and holdings of asset A is achieved.*

The above examples convey how remarkable it is that one participant of either type trade to the market. We do not see many participants that do this, but we do see the average and median participant having the same state-holdings as the market. Moreover, participant state-holdings in incomplete market periods are indistinguishable from their holdings in complete market periods. These findings are in agreement with those of prior asset pricing experiments, and are compatible with correct state-price ranking within the framework of the CAPM+ϵ model.

We display relative state-holdings for each announcement for all relevant periods, side by side in Figures 2.10 to 2.13. The figures support the findings mentioned above.

2.5 Conclusion

We have shown that state-price probability ranking according to theory obtains in experimental dynamically complete markets. Moreover, the data support the claim that this ranking obtains because participants are able to form the right expectations about future prices and the median participant trades towards the market portfolio.

It is important to stress that we design our experiments such that the above results obtain in an environment where participants can hold the market portfolio only if they re-trade after the announcement that completes the markets. Careful design of the experiment allowed us to recover state-price probabilities from this environment where the theory is put to a very hard test.

Correct state-price probability ranking in the presence of median and mean trade towards the market portfolio is a result of the CAPM+ϵ model proposed in Bossaerts, Plott, and Zame [2005] for a complete market environment. In view of our experimental results it is important to return to the theoretical framework and develop the dynamical structure necessary to evaluate the model with our data for dynamically complete markets.

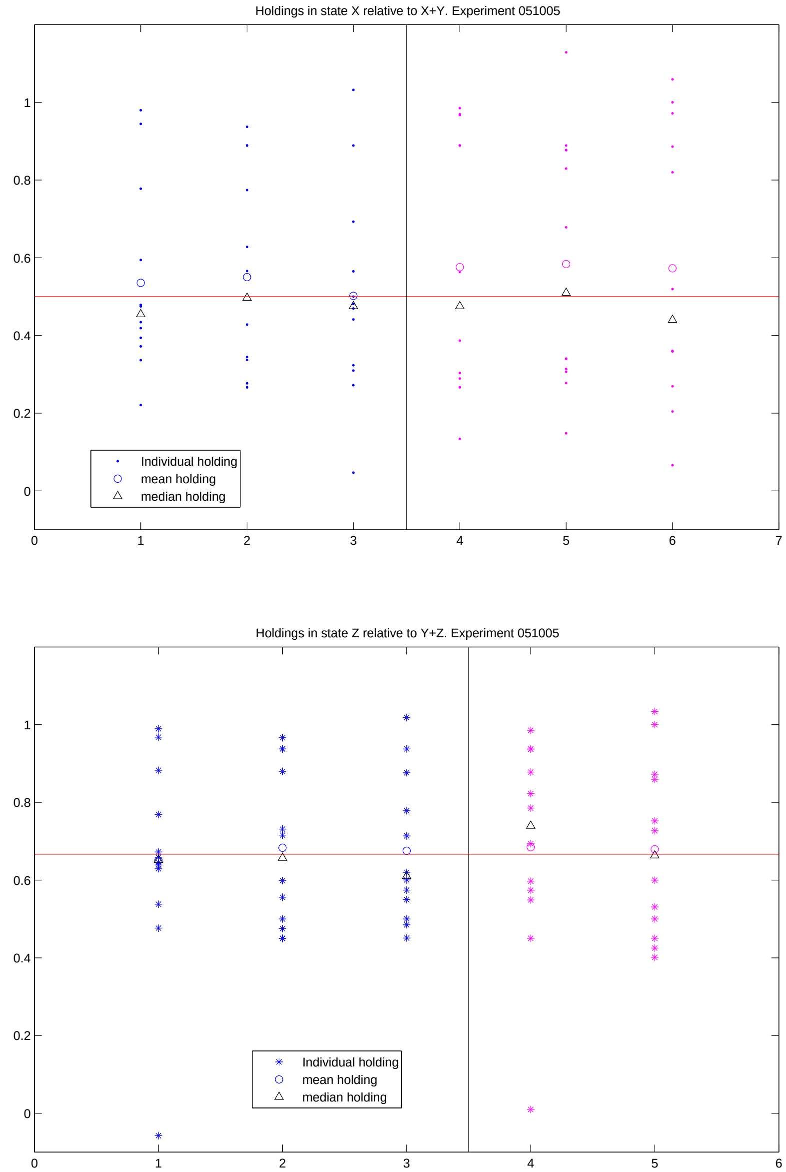


Figure 2.10. Relative state holdings for periods with announcement “not Z” and announcement “not X”. The same relative state holdings are computed for complete market periods and reported in the same figure for comparison. Periods left of the vertical line are complete market periods. The horizontal line represents the relative state holding given by the market portfolio.

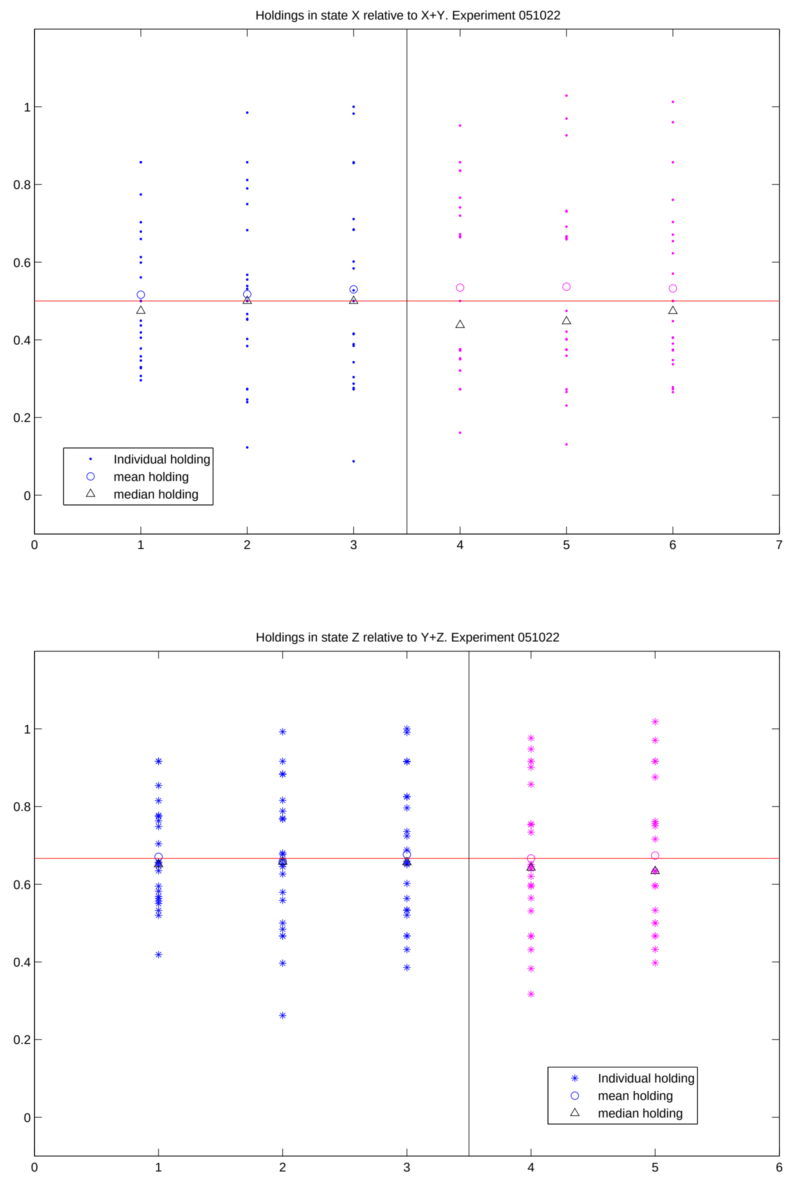
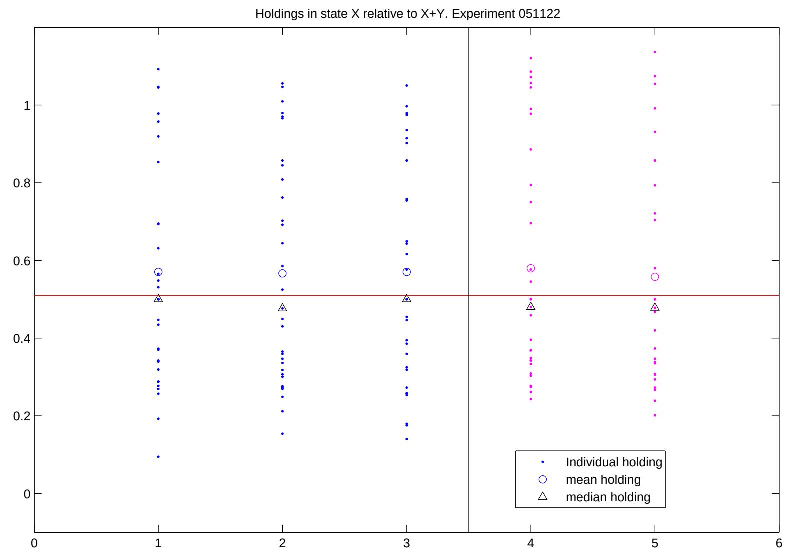


Figure 2.11. Relative state holdings for periods with announcement "not Z" and announcement "not X". The same relative state holdings are computed for complete market periods and reported in the same figure for comparison. Periods left of the vertical line are complete market periods. The horizontal line represents the relative state holding given by the market portfolio.



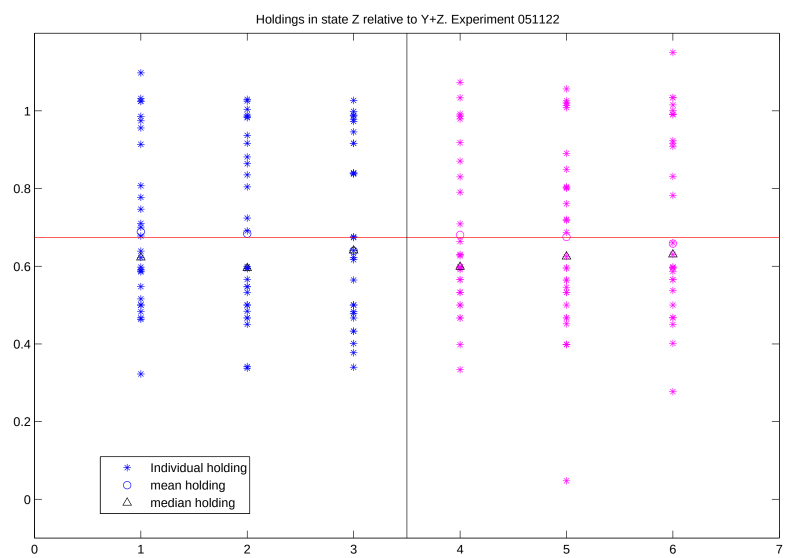
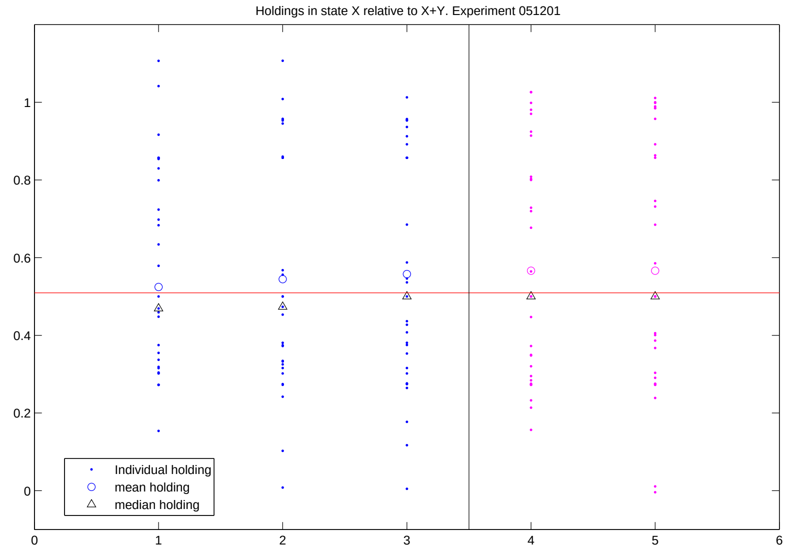


Figure 2.12. Relative state holdings for periods with announcement "not Z" and announcement "not X". The same relative state holdings are computed for complete market periods and reported in the same figure for comparison. Periods left of the vertical line are complete market periods. The horizontal line represents the relative state holding given by the market portfolio.



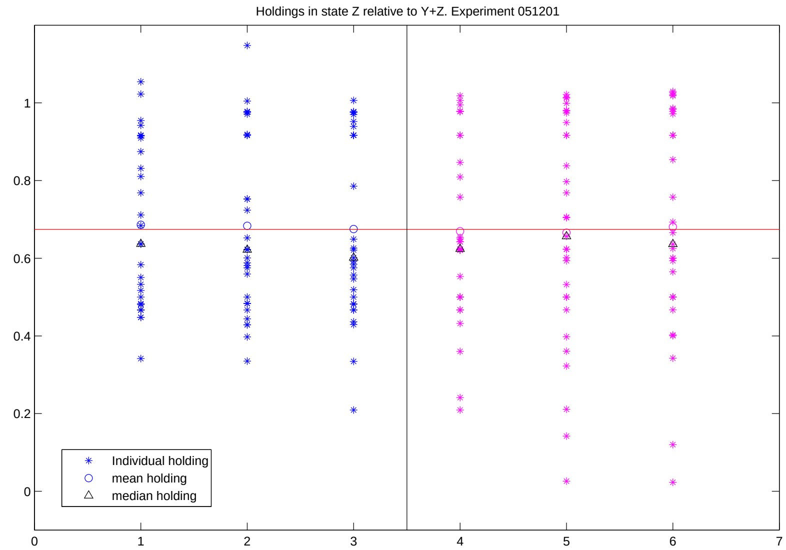


Figure 2.13. Relative state holdings for periods with announcement "not Z" and announcement "not X". The same relative state holdings are computed for complete market periods and reported in the same figure for comparison. Periods left of the vertical line are complete market periods. The horizontal line represents the relative state holding given by the market portfolio.

Chapter 3

Executing Complex Cognitive Tasks: Prizes vs. Markets

Abstract

Execution of complex cognitive tasks is often analyzed as an exercise of information acquisition and belief updating. We challenge this view in the context of a non-incremental task, namely, the knapsack problem. First, we provide a theoretical argument why Bayesian updating makes little sense in this context. Second, we provide experimental evidence against the Bayesian approach by comparing the quality of problem solving under two treatments: prizes; markets. We find that Bayesian theory cannot make sense of the data: both systems work equally well, while trading is abundant in the market setup and prices are informative but noisy. The experimental data provide suggestions for a new theory of discovery of solutions in non-incremental tasks.

3.1 Introduction

In the standard economic model, whatever is not known is treated as a random variable over which one holds a prior belief. Learning and discovery are modeled as an application of Bayes' law to construct a posterior.[1] Though this may be close to how we think when we try to guess the number of coins in a large closed jar, it is not very descriptive of how Picasso thought of "Guernica." Here we explore this latter form of discovery. We run experiments with this purpose, and also to demonstrate that consideration of the characteristics of the discovery problem at hand, matters in economically relevant situations.

The creation of a piece of art like Guernica is not readily achieved in the laboratory. This is why we resort to a cognitive task that captures the features of discovery that we care about - mainly, that it cannot readily be translated into Bayesian learning. The cognitive task we consider is the knapsack problem. An instance of the knapsack problem is given by a set of indivisible objects defined by a value and a weight parameter. The objective is to find the combination of objects that produces the largest sum of their values, given that the sum of their weights is within a given limit. The knapsack problem suits our objective because it is non-incremental and hard (we will return to these ideas below). We show that, as an implication of non-incrementality, the Bayesian paradigm does not provide a reasonable model of the search for a solution in an instance of the knapsack problem.

Our experiment then serves two purposes. First, it demonstrates that it is relevant to Economics that discovery may be non-incremental and, thus, existing models do not address the implications of this notion. Second, it points out in what ways non-incrementality differs from traditional approaches, thus opening the path for further experimental and theoretical exploration of the topic.

In our experiment we look at how institutions interact with the task of solving an instance of the knapsack problem. We consider two institutions which we refer to as treatments; a prize and a market treatment. For traditional, Bayesian problems, where discovery is treated as updating of information, the market treatment provides very little incentive for discovery (see, e.g., Grossman and Stiglitz [1980]). On the other hand, we calibrate the two treatments in such a way that the prize to one participant in the prize treatment equals what we pay to the market as a whole in the market treatment, thus providing a large individual incentive conducive to discovery. We find that markets do very well, according to whether somebody finds the solution, and how many persons find it. We are also able to relate the performance of markets with the trading behavior of participants and the pricing of securities.

The feature of the knapsack problem we care most about is non-incrementality. This means that many instances cannot be divided into a small number of parts that can easily be computed, finding the solution of the entire instance through computation of these parts. The knapsack problem is also hard; even though many instances can be solved using heuristics, it is typically not possible to determine before the problem has been solved, what heuristic will work for the instance at hand. Known exact solution algorithms that

[1]Examples include probabilistic learning in markets (Grossman [1977]); or how individuals learn their own productivity (Jovanovic [1979]). A reduced form of this approach is common in the literature on Intellectual Discovery (see Arrow [62], and Gallini and Scotchmer [2002] for a survey). Commentary and a brief review of the classical study of technological change is given in Arrow [1994].

will always work are laborious, and the time they take rapidly increases with the size of the instance. Both features relate to the fact that no polynomial time algorithm has been found to solve every instance of the knapsack problem.[2] This feature is used in section 3.2.3 to show that Bayesian updating is not reasonable in our setting. Loosely speaking, the argument is that in hard-enough instances of the knapsack problem, failure to compute the solution stems partly from the size of the set of possible solutions. This implies that if a subject were not able to compute the optimal solution, then he would also not be able to make any useful updating of his beliefs over the set of possible solutions. Thus, instances of knapsack problems can be constructed such that either there will be no uncertainty (if subjects can compute the solution fast) or subjects will not be able to compute the solution, but the Bayesian paradigm is then also inadequate.

The fact that the knapsack problem is non-incremental affects the possibility of coordination and decentralized computation in the markets we implement. In particular, assigning to different participants the task of finding the value of one security leads to no gain at all in the ease of computation. This is so because our securities correspond to objects in the knapsack problem, each paying a positive dividend only if the corresponding object belongs to the value-maximizing knapsack. This splits up the instance in a very small number of pieces, in which case non-incrementality implies that each piece must be hard to compute (this depends on the instance, as we already remarked). In fact, for the instances we use, for all objects, the problem of determining whether an object belongs to the optimal knapsack or not, is equivalent to finding the optimal knapsack. Thus, even though the securities in the market provide a language that may serve to coordinate decentralized computation of the solution, they do not correspond to tasks for decentralization. It is important to note that distributed computation is not the only possible source of gains from interaction among individuals confronted with a cognitive task. This is suggested by literature in Psychology. For example Maciejovsky and Budescu [2005] reports an extensive experiment on learning in groups, where they find that groups are able to correct individual biases (mistakes) even when groups are composed only of individuals who display these biases when alone.[3]

Within the Bayesian, information aggregation framework, the securities in our market treatment are expected to lead to poor communication and aggregation. This is so because they do not allow for the expression of conditional beliefs like, for example, "conditional on object A being in the optimal knapsack, object C has probability 0.3 of also being there." Ledyard [2005] shows that this absence significantly decreases the informational quality of the market posterior. In a similar setup, Plott and Sunder [1988] shows that a complete set of contingent securities yields better information aggregation than a single security with multiple payoff levels.[4] We thus reiterate that our market treatment is not an attempt to implement

[2] There are other problems that satisfy these properties. We choose the knapsack problem because it is well suited for implementing the market treatment. One example of a non-incremental problem is that of finding the "right" linear regression, as pointed out in Aragones et al. [2005]. The findings in this work are also a good example of how intellectual discovery or learning has incremental (running the regression) as well as non-incremental (deciding what regression to run) parts. For an extensive discussion on the incremental and non-incremental features of research and science, see Kuhn [1962]

[3] The gains or losses from interaction have also been explored in Economics. Blinder and Morgan [2005] report that groups make better decisions than individuals conditional on equal amounts of information, while Cox and Haynes [2006] finds that groups fall victim to the winner's curse in auctions more often than individuals.

[4] We say "similar" because the single security in Plott and Sunder [1988] is more akin to, for example, having only one security paying the value of the optimal knapsack in our setup.

the most efficient mechanism for *information aggregation*, but rather a tool to understand non-incremental discovery within the frame of economic institutions.

This paper points out that the description of intellectual discovery is relevant to our understanding of incentives to foster it. It also points out that cognition and computation may be the driving forces of trading and pricing behavior in complex information aggregation setups (e.g., Ledyard [2005]). As noted before, we also wish to better understand non-incremental discovery. To this end, we construct two ad hoc measures of difficulty of an instance, design our instances to have a range of difficulties according to these measures, and look for correlation between experimental results and the measures we propose. We find that the difficulty of the approximation algorithm - within a narrow class - that exactly solves an instance, is a good predictor of the number of individuals that solve the instance. In section 3.2.2 we explain the two measures we propose and discuss why our results are surprising.

The remainder of the paper is organized as follows: In section 3.2 we extensively describe the knapsack problem. In Subsection 3.2.3 we present an illuminating result about the knapsack problem being outside the scope of Bayesian updating. Section 3.3 presents and discusses the experimental design, with its different treatments. Section 3.4 presents the results. Results and future research are discussed in section 3.5, which concludes.

3.2 The Knapsack Problem

The Knapsack Problem represents the model of intellectual discovery we wish to deal with. In this section we will carefully define the knapsack problem and some of its properties.

An instance of the knapsack problem is defined by a set N, containing n objects (we will refer to n as the *size* of the instance of the knapsack problem) with values $v = (v_1, \ldots, v_n)$, and weights $w = (w_1, \ldots, w_n)$; and a weight limit c. In the knapsack problem, the objective is to find the subset of N that yields the maximal sum of object values, given that the sum of weights does not exceed c. We will call this subset the *optimal knapsack*. The above objective can be stated mathematically as:

$$\begin{aligned}
&\max \sum_{j=1}^{n} v_j \theta_j \\
&\text{s.t. } \sum_{j=1}^{n} w_j \theta_j \leq c \\
&\theta_j \in \{0,1\}.
\end{aligned}$$

An intuitive but incorrect approach to solving the knapsack problem is the *greedy algorithm*. According to this algorithm, the object with highest value to weight ratio, $\frac{v_j}{w_j}$ (the most *efficient* object) is added first, followed by the object with the second highest value to weight ratio, and so on, until the weight limit is hit. This procedure can yield an arbitrarily bad approximation to the true solution (measured in terms of the value of the knapsack).

There are no known simple algorithms that will solve every instance of the knapsack problem. On the other hand, the related linear programming problem,

$$\begin{aligned} &\max \sum_{j=1}^{n} v_j \theta_j \\ &\text{s.t. } \sum_{j=1}^{n} w_j \theta_j \leq c \\ &\theta_j \in [0,1], \end{aligned}$$

has a straightforward solution, established by Dantzig [1957] that can always be found in linear time in the size of the instance. The solution of the above linear program is found by first applying the greedy algorithm. The first object that is left out according to the greedy algorithm, is then split so to fill the remainder weight.[5]

The following example serves to fixate notation and illustrate the working of the greedy algorithm:

Example 7. *Here is an instance of the knapsack problem:*

$$v = \begin{bmatrix} 5 \\ 7 \\ 11 \\ 20 \end{bmatrix}, w = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6 \end{bmatrix}, c = 7.$$

There are four objects ($n = 4$)*, their values are in* v*, and their weights in* w*, the weight limit is* $c = 7$*.*

Use of the greedy algorithm yields the following solution for the instance we consider here:

$$\hat{\theta} = (1, 1, 1, 0).$$

Knapsack $\hat{\theta}$ *contains the first three objects, has a total weight of* 6*, and a total value of* 23*. However,* $\hat{\theta}$ *is not the* optimal *knapsack, which illustrates that the greedy algorithm is not a correct solution method. Instead, the* optimal *knapsack is given by*

$$\theta^* = (1, 0, 0, 1),$$

containing objects 1 *and* 4*, and with a total weight of* 7*, and a total value of* 25*.*

3.2.1 NP-Completeness of the Knapsack Problem

A problem is said to be solvable in polynomial time if there exists an algorithm that solves every instance in a running time that is bounded above by a polynomial function of the instance size, n. A problem Q is

[5] There are many references for more on the knapsack problem, its variants, bounds, and exact and approximate algorithms. Two excellent guides are Kellerer et al. [2004] and Martello and Toth [1990].

$\mathcal{NP}$-complete ($\mathcal{NPC}$) if it is $\mathcal{NP}$ and $\mathcal{NP}$-hard. The former means that a proposed solution to Q can be *verified* in polynomial time, and the latter means that every problem in the class $\mathcal{NP}$ can be transformed into Q in polynomial time. No algorithm running in polynomial time in the instance size is known for any problem in the $\mathcal{NPC}$ class. If such an algorithm were found for one problem in $\mathcal{NPC}$, then this would immediately establish that all problems in $\mathcal{NPC}$ belong to $\mathcal{P}$, the set of problems that can be solved in polynomial time.

To us it is most relevant that no polynomial-time algorithm is known for the knapsack problem. This means that no fast (clever) algorithm can be used to solve every instance. In particular, if a participant in our experiment solves one instance, this does not mean that he has *learned* to solve the problem and will henceforth face no more difficulties.[6] However, it does not mean that every instance of the knapsack problem is difficult. First, if the instance is sufficiently small, the crudest exact solution algorithm will still solve it in a fairly small number of computations (e.g., the largest instance used in our experiment has 12 objects, which implies that the value of only 4096 knapsacks must be computed and compared to find the solution). Second, and more important, certain instances can be exactly solved with fast approximate algorithms, e.g. the greedy algorithm. This is the subject of the next subsection.

3.2.2 Measures of Difficulty of Instances

There is no straightforward measure of the difficulty of an instance of the knapsack problem. For example, even though in the *worst case* an instance of size $n + 1$ may take twice as much time as one of size n, this pattern need not be satisfied by two arbitrary instances of size $n + 1$ and n, respectively. We propose two measures of difficulty, and revisit them when we present our choice of instances for the experiments we ran.

A first intuitive measure is a parameter of *input size*. It is the product of the base-two logarithm of the knapsack capacity times the size of the instance. The base-two logarithm of the knapsack capacity is a proxy for the binary representation of the instance parameters, which in turn represents the amount of storage and information necessary in each step of computation. This is then multiplied by the size (the number of objects) of the instance. In this measure, two instances with equal capacity and size are equally difficult. We propose a second measure which captures less obvious characteristics of an instance.

This second measure relates to the question of heuristic solvability of an instance.[7] A simple approximation algorithm for the knapsack problem may solve an instance exactly. For a class of approximation algorithms described below, we will consider an instance that can be exactly solved with an approximation algorithm in this class to be easier than another instance that cannot be exactly solved with it.

The simplest approximation algorithm for the knapsack problem is the *greedy* procedure, which consists of filling the knapsack in efficiency order - i.e., starting with the objects that have a higher $\frac{v_i}{w_i}$ ratio - until

[6]One implication of this is that the periods in our experiment are independent instances of discovery, in the sense that having solved one instance does not guarantee a participant will solve every other instance. Still, participants' skills may improve. For example, a participant that starts by using the greedy algorithm, will quickly learn it is not a very good approach, or the clumsy exact algorithm that explores every possible knapsack may be replaced with a more efficient branch-and-bound algorithm.

[7]We use the terms *heuristic* and *approximation algorithm* interchangeably.

the weight limit is reached. This heuristic is part of a family of approximation algorithms known as the Sahni approximation scheme (Sahni [1975]).[8] The Sahni scheme is parameterized by a number k, referring to the specific algorithm, which we will call a *Sahni algorithm of size k*.

A Sahni algorithm of size k looks at all subsets of N (the set of all objects that are considered to enter the knapsack) of cardinality k or less. For each subset, it computes the residual weight in the knapsack after subtracting the weight of the subset, and fills this residual with the remaining objects using the greedy procedure (if no set of k objects fits in the knapsack, then the exact optimal solution is found by the algorithm). The value of all knapsacks constructed in this way is compared, and the one with highest value becomes the approximate solution given by the algorithm. Clearly, the greedy algorithm is a Sahni algorithm of size 0. The Sahni algorithm of size 1 uses the greedy algorithm for every subset of size $n-1$ (there are n such subsets) to fill the capacity that remains after isolating one object. Though the complexity of the greedy algorithm it runs is smaller, it has to run it n times. The complexity added is thus of order n. This is the case for every increase from k to $k+1$. Although a Sahni algorithm of size k is not as straightforward as the greedy algorithm, it is still a very simple heuristic.

Definition 7. *We say that an instance has Sahni-difficulty level k if it can be exactly solved with a Sahni algorithm of size k, but not with a Sahni algorithm of size $k-1$. The higher the k associated to an instance, the harder the instance.*

While specific instances can be solved using simple heuristics, it is impossible to determine a-priori what simple heuristic to use. Given an instance, if one knew a priori what Sahni algorithm to use in order to find the solution, then the Sahni-k would be a precise measure of difficulty. However, only after the instance has been solved can one know what Sahni algorithm to use. Thus, there is no reason to believe that Sahni-k will be a good predictor of whether a person can solve an instance or not.

In table 3.2 we return to the two measures of difficulty and apply them to the instances used in our experimental sessions.

3.2.3 The Knapsack Problem: A Hard Problem

We wish to draw a parallel between non-incremental intellectual discovery and finding the solution to a hard problem. The knapsack problem allows for instances that are hard according to our criterion because of the following properties:

1. There is a parameter (in our case the size, n), such that an increase of this parameter by 1 unit can in some instances increase the computational complexity of finding the exact solution by an order of magnitude. This allows us to construct problems that are very hard.

2. It is not possible to establish a priori what approximation algorithm may solve the instance exactly or

[8] Approximation schemes have more desirable properties than approximation algorithms. In the case of the Sahni scheme, different algorithms from the family can be chosen depending on the desired performance level.

with a fixed precision. It is also not possible to split the problem in a *small number* of *easy* parts such that solving each of these parts separately will give the right answer.

The above two properties make the problem *non-incremental.* Additionally, the knapsack problem satisfies the following property, which allows us to prove the main result of this section:

3. When n is increased by 1 unit, the complexity of computation of the value function of the instance at each point in the solution space does not increase very much.

When it comes to identifying a hard problem, we just care that we can find *an instance* of the knapsack problem which is hard enough for our purposes and which satisfies property 3. This contrasts the worst-case analysis that was mentioned in section 3.2.1, and which is often used in computer science.

The reason we care about properties 1, 2, and 3 is that these properties allow us to make a clear-cut distinction between our setup and rational-expectations models. In rough strokes, a problem that satisfies properties 1, 2, and 3, will either be easy enough to be computable by agents, or it will be too hard to be computable. In the latter case, it will *also be too hard to update any prior belief over the solution set.* Hence, in our model, there is either no uncertainty, or agents cannot possibly update any prior over the solution set, so that any model involving updating of beliefs cannot be the right one to model this situation. This provides the basis for testing the hypothesis of whether we posed "hard enough instances" of the knapsack problem to our subjects. We now make our argument precise.

Take the following simple Bayesian model to describe solving the knapsack problem as a situation with uncertainty in which agents get new signals and are required to update their beliefs over how likely each point in the solution set is to be the optimum. Call this model the *Uncertainty Model.*

Let $\Theta(n)$ be the set of *possible* solutions of an instance of the knapsack problem with n objects. Thus, $\Theta(n)$ is in the case of the knapsack problem the set 2^n. Each agent $i \in I$ (I is some non-empty index set) has a prior belief B_i over $\Theta(n)$ describing for each $\theta \in \Theta(n)$ how likely it is that θ is the solution to the specific instance of the problem. Assume that $B_i(\theta) > 0, \forall \theta \in \Theta(N)$. This is a sensible assumption since the complexity of the problem is on one hand determined by the number of *possible* solutions, so that points with 0 probability do not add any complexity.[9] Now assume that i observes a realization of a random variable $\sigma \in \{0, 1\}$, and there is an updating mapping $B_i(\theta \mid \sigma) : \{0, 1\} \times B_i(\theta) \rightarrow [0, 1]$, specifying how the probability of each state changes depending on the realization of σ. We assume that this updating does not reduce the complexity of the problem by an order of magnitude. If this were not the case, and if we want to think of σ as some accessible and simple device available to the agent, then he could just use n such devices and figure out the precise solution in polynomial time. That such a scheme is not readily available to the agent is therefore an assumption consistent with the problem being $\mathcal{NP}$.

[9] This assumption is additionally justified as follows. For example, if all objects have a small w_j and the weight limit is large, one expects an individual to immediately rule out all knapsacks with only one object. One should think of this as a signal, as this can only be established once an agent is presented with the instance; the updated distribution gives probability zero to all knapsacks with only one object. This signal satisfies the assumption we make about signals. The assumption on prior belief can also be replaced with the alternative assumption that the support of $B_i(\theta)$ be of cardinality larger than 2^{n-1}, or in fact with any cardinality that is exponential in n.

Finally, we assume that the complexity of computation of the updated probability $B_i(\theta \mid \sigma)$ is of the same order of magnitude as the computation of the value function $V(.)$ of the instance of the problem at each θ. This assumption is sensible precisely because of 1–3. Namely, suppose that we increase n by 1 unit to n'. Then the set $\Theta(n')$ increases in size by a factor of 2 relative to $\Theta(n)$, or an *order of magnitude*. However, computing $V(\theta)$ either remains unchanged in complexity for $\theta \in \Theta(n) \cap \Theta(n')$, or it at most increases by an additive constant for $\theta \in \Theta(n') \setminus \Theta(n)$.[10] Formally, for each $\theta \in \Theta(n)$ denote by $\eta_{i,B}(\theta, n, \sigma)$ the number of operations needed to compute $B_i(\theta \mid \sigma)$, and denote by $\eta_V(\theta, n)$ the number of operations needed to compute $V(\omega, n)$. To avoid any confusion with computer science, we state our assumption as follows:[11]

$$\frac{\eta_{i,B}(\theta, n, \sigma)}{\eta_V(\theta, n)} \geq \underline{m} > 0, \forall \theta, \forall n \leq \bar{n} < \infty.$$

Here, $\underline{m}$ is a positive constant, *close to* 1, and $\bar{n}$ is some potentially very large constant. It is clear that we can restrict ourselves to finite problems, as when $\bar{n}$ is for instance 2 billion, the solution set of the knapsack problem with so many objects is much larger than the number of atoms in the universe, and there is clearly no sense in theorizing about problems that complex.

To make things more concrete and directly related to our problem, think of σ as a comparison of the values of two different knapsacks. Such a σ is a very natural one, and it clearly satisfies the above assumptions.

Proposition 8. *In the* Uncertainty Model*, assume that* $\underline{m} \geq 1$*. Then, if agent* i *is unable to compute the solution to the instance of the knapsack problem,* $\theta^* \in \Theta(n)$*, then it is also for* i *impossible to compute* $B_i(. \mid \sigma)$*.*

Proof. The proof is obvious: if i were able to compute $B_i(\theta \mid \sigma)$, for all $\theta \in \Theta(n)$, then he could have just as well computed $V(\theta)$ at each $\theta \in \Theta(N)$ in the first place, and then keeping always track of just the maximal θ he would find the precise solution θ^* so that there could be no uncertainty.

We comment that $\underline{m} \geq 1$ is just a convenient device to make our statement particularly simple, and the mapping $B_i(. \mid \sigma)$ is abstract and might be very complex anyway. Even if $\underline{m}$ were equal to $\frac{1}{n}$,[12] a similar statement would apply: the only difference is that then, if the agent were unable to compute the solution for a specific n, he surely would be unable to compute the posterior when n is increased by 1 unit.

[10] Computing the value of the knapsack is just adding up the values of all the objects, and adding all the weights to verify that the weight limit is not surpassed. Clearly, if we have one new object there is at most one new weight and one new value to add, so that computational complexity can at most increase by 2 operations. Moreover, even at most points $\theta \in \Theta(n') \setminus \Theta(n)$ computing the value does not increase in complexity at all.

[11] We could state our assumption in computer-science language as $\eta_V(\theta, n) = O\left(\eta_{i,B}(\theta, n, \sigma)\right)$, but such a statement is one of asymptotic properties where n becomes large. Our statement describes the situation for finite n.

[12] This is the worst things could possibly go for us, as the complexity of computing $B_i(. \mid \sigma; n)$ is at least a constant and the complexity of computing $V(\theta, n)$ is at most of the order n. The constant in the numerator is 1 because all values of the parameters in an instance have the same base-two log.

3.3 Experimental Design

The instruction sets can be found by visiting the following web pages (which were part of the web site that was used in the experiments):

http://clef.caltech.edu/exp/eTradeLab13a/instructions.html and
http://clef.caltech.edu/exp/knapsack.

The first webpage corresponds to the original experiment, of which we ran four sessions. We will refer to this setup as setup ω. The second webpage corresponds to the version with equal time for both treatments and a variation in the prize setup that is explained below. We will call this setup, setup ν.

3.3.1 Timing of an Experimental Session

In an *experimental session*, a group of participants are asked to solve several instances of the knapsack problem, cast in terms of *shipping problems*. All instances have a unique solution. Each experimental session is divided in ten *periods*, $P = 1, ..., 10$. Periods 1 and 2 are *practice* periods; the remaining ones (3 to 10) are *earnings* periods. The explanation that follows applies to both setups - ω and ν - except when specified.

In setup ω, prizes, security prices, cash, and period earnings are expressed in a fictitious currency called *Franc* (F), with an exchange rate of $F1 = \$0.01$. In setup ν, the dollar value of all parameters is preserved, but expressed in dollars. In both setups, earnings from practice periods do not count towards total experiment payoffs. Earnings from earnings periods are accumulated, converted to dollars at a preannounced exchange rate, and a $5 sign-up reward is added. In each earnings period, participants are given an instance of the knapsack problem and a specific payment rule. Under the *Market* treatment, participants are paid based on their final holdings of securities whose dividends depend on the optimal solution, as well as their final cash position. Under the *Prize* treatment, they are paid directly for solving the instance of the knapsack problem. All odd-numbered earnings periods are *Market* periods; all even-numbered earnings periods are *Prize* periods.

At the beginning of a period, participants receive a sheet of paper with a full description of an instance of the knapsack problem and an area to mark their proposed solution to the problem. This sheet will be referred to as *Answer Sheet* (see *Shipping Problem Presentation* in the Instructions Sets). Answer Sheets are collected at the end of a *Prize* period; in a *Market* period, the sheets are collected 30 seconds before the end. In other words, the Answer Sheets are handed in by participants, before security payoffs are revealed, and hence, before the optimal solution, which is implicit in the security payoffs, becomes known.

An entire experimental session lasts approximately 2 1/2 hours, including one half hour for instructions. In setup ω, each prize period lasts up to 7 minutes, and each market period lasts 15 minutes. In setup ν, each period lasts 10 minutes, for both treatments.

Table 3.1. Instances of the knapsack problem used in experimental sessions. Objects had common names, not letters. The letters used in the table stand for: $A = Anderson$, $B = Brown$, $C = Cole$, $D = Darwin$, $E = Evans$, $F = Foster$, $G = Green$, $H = Hamilton$, $I = Ives$, $J = Jensen$, $K = Keaton$, $L = Lee$.

Problem & capacity		Objects											
		A	B	C	D	E	F	G	H	I	J	K	L
I	p	500	350	505	505	640	435	465	50	220	170		
	w	750	406	564	595	803	489	641	177	330	252		
$c = 1900$	θ^*	0	0	1	1	0	1	0	0	0	1		
II	p	300	350	400	450	47	20	8	70	5	5		
	w	205	252	352	447	114	50	28	251	19	20		
$c = 1044$	θ^*	1	0	1	1	0	0	0	0	1	1		
III	p	15	14	3	3	10	9	28	28	31	25	24	1
	w	129	144	77	77	66	60	184	184	229	184	219	72
$c = 850$	θ^*	0	0	0	0	1	0	1	1	1	1	0	0
IV	p	37	72	106	32	45	71	23	44	85	62		
	w	50	820	700	46	220	530	107	180	435	360		
$c = 1500$	θ^*	1	0	0	1	1	0	1	1	1	1		
V	p	2	3	4	5	6	9	8	7	6	5	8	9
	w	3	4	6	3	5	13	6	9	2	4	7	7
$c = 14$	θ^*	0	0	0	1	1	0	0	0	1	1	0	0
VI	p	107	35	120	206	88	34	28	110	88	101	74	53
	w	599	196	670	1204	502	202	145	600	453	601	404	299
$c = 3800$	θ^*	1	1	0	0	1	0	1	1	1	1	1	1
VII	p	201	84	113	303	227	251	129	147	86	127	144	167
	w	192	80	106	288	212	240	121	140	82	120	137	160
$c = 1300$	θ^*	1	0	1	1	1	0	0	1	1	1	0	1
VIII	p	31	141	46	30	74	105	119	160	59	71		
	w	21	97	32	21	52	75	86	116	43	54		
$c = 265$	θ^*	0	1	0	0	1	0	0	1	0	0		

3.3.2 Knapsack Instances

Table 3.1 lists the instances of the knapsack that were used in the experiment. There are eight different instances; four have n (number of items) equal to 10, and four have n equal to 12.

In each experimental session, two different instances of size $n = 10$ are assigned under the Market treatment, and two different ones are assigned under the Prize treatment. The analogous statement is true for instances of size 12. To ensure that all instances are solved under both setups, we run two types of sessions. In experimental sessions of type a, instances I, V, VII, and VIII (see table 3.1) are solved under the Market treatment; while instances II, III, IV, and VI are solved under the Prize treatment. In experimental sessions of type b, this organization is reversed: an instance that is solved under the Market treatment in type a is solved under the Prize treatment in type b, and vice versa.

3.3.3 Market Treatment

In the Market treatment, participants are paid through securities. There are as many securities as there are items in the knapsack instance. Each security corresponds to an item. At the end of the period, a security pays \$1 if the corresponding object is in the optimal knapsack; otherwise the security pays nothing.

Participants start with an initial endowment of 5 units of each security, and \$4 cash (this is expressed as $F400$ in setup ω). To provide liquidity, participants are endowed with more securities than the desired average payoff; this is why a *"loan repayment"* is subtracted from total period earnings. In type a experiments the loan repayment is \$23.75; in type b experiments the loan repayment is \$32.50. Because of the loan, participants may lose money. Losses are subtracted from cumulative earnings. If cumulative earnings are negative, the participant is paid only the sign-up reward of \$5 plus earnings from prize periods.

In setup ω, trading is done through a continuous electronic open-book system called eTradeLab. Details of this trading interface can be found in the instructions (see aforementioned webpage). The system tracks all offers (bids, asks) and transactions, time stamped to the second.

In setup ν, an identical market mechanism is used for trading. However, the interface is different, and orders are submitted by "clicking" on price labels, instead of manually entering a price and quantity. The software for this setup is jMarkets, and is further described in the experiment instructions for setup ν.

While Answer Sheets are collected 30 seconds before the end of the period (i.e., before security payoffs are revealed), participants are not paid depending on their marks on these sheets. Participant earnings for Market periods come exclusively from trading and from the final payoffs of the securities.

3.3.4 Prize Treatment

The prize treatment in setup ω is substantially different from that in setup ν. We describe them separately, under the names *prize treatment* ω and *prize treatment* ν, respectively.

3.3.4.1 Prize Treatment ω

At any moment during a Prize period, a participant can submit his/her Answer Sheet with the proposed solution by raising his/her hand. The experimenter then checks the marks on the Answer Sheet and announces a winner if the solution is correct. The period ends when all participants have submitted their Answer Sheets, or the time limit has been reached (recall that the time limit is 7 minutes), whichever occurs first. Participants do not get a second chance: once an Answer Sheet is turned in, participants cannot change it. The Prize is set at $F6,600$, i.e., \$66. Ties are resolved by dividing the Prize equally among the winners.

3.3.4.2 Prize Treatment ν

All participants are in constant live communication with the experimenter, through a chat program called Skype. A username and password are assigned to each participant for the duration of the experiment. These are secret, to ensure that participants communicate only with the experimenter. Participants have one

Table 3.2. Difficulty of Knapsack Problem instances used in experiments. The difficulty is measured by the proxy of *input size* (columns) and the Sahni difficulty level described in section 3.2.2.

		Input Size Proxy ($n \log_2 c$)				
		40 - 50	80 - 90	100 - 120	120 - 130	140 - 150
Sahni difficulty level	0			IV		
	1	V		I		
	2		VIII	III		
	3			II		VI
	6				VII	

opportunity to submit their proposed solution to the instance at hand. The submission is made over the chat program, using a code of ones and zeros, to indicate objects that are in and objects that are not in the optimal knapsack according to their proposed solution. Participants have a time limit of 10 minutes to make submissions. After these 10 minutes elapse, all answer sheets are picked up, after which the correct answer is announced together with the time stamp of the winning submission. The winner is the first participant to submit the correct solution. Unlike in setup ω, in prize treatment ν, the fact that there is a winner is not revealed until the end of the period.

Note that in both setups the Prize is approximately the same as the *aggregate* payment in Market periods (security dividends and cash, minus loans). Notice the difference between a Prize period and a Market period: in a Prize period, only the winner(s) is (are) paid; in a Market period, everyone is paid, through securities and cash.

3.3.5 Discussion of Experimental Design

3.3.5.1 Difficulty of Knapsack Problem Instances

The instances we consider have different levels of difficulty according to the two measures given in subsection 3.2.2. In table 3.2 below, the rows correspond to Sahni difficulty levels, k, while the columns correspond to the described parameter of *input size*. In table 3.2, the northwest corner corresponds to lower complexity. Complexity increases in the southeast direction.

3.3.5.2 All-or-Nothing Problem

One property of the knapsack problem is that a knapsack that is close to optimal - say, the feasible knapsack with the second-highest value - need not look in any way similar to the optimal knapsack. The latter may be composed of an entirely different set of objects. We want to preserve this property in our experimental treatments. It is to this end that compensation is never tied to the value of the knapsack that a participant proposes as the optimal knapsack.

It is obvious that compensation in the prize treatment is given only for the exact solution of the problem. We wish to emphasize that this is also the case in the market treatment. Only securities corresponding to objects that belong to the optimal knapsack pay positive dividends. For example, if the knapsack with the second-highest value has an entirely different composition from the optimal knapsack, a participant betting on the objects composing the former will make negative earnings. In section 3.5 we further discuss the payoff structure in our experiment.

3.3.5.3 Securities Do Not Split the Problem

As discussed in section 3.2.3, instances of the knapsack problem satisfy property 2, i.e., the solution cannot be found by splitting the problem in a small number of easily-solvable parts. In particular, computing whether an object belongs to the optimal knapsack or not is as hard as solving the instance.[13] Therefore, computation of the solution cannot be decentralized in an obvious fashion by each subject focusing on a specific object.

One example of how the process of finding the solution to an instance can be distributed is by splitting the set of all knapsacks in groups for comparison. Each person gets a fraction of all knapsacks to make value comparisons. The most valuable feasible knapsack of each group is then compared to the winner of other groups, and the most valuable feasible knapsack can thus be found.[14] It is not clear how this form of distribution can be encoded in the securities in the market treatment.

The securities in the market are the minimal set of securities necessary to express the solution of every instance. They provide a binary code to represent the solution, and may serve as a language to transmit it. However, the securities do not distribute or simplify the problem in any way.

3.4 Results

We report results from four ω experimental sessions and four ν sessions. Sessions are identified by the date when they were run (yymmdd), the setup (ω or ν), and the experiment type (a or b). Our eight sessions are: $\omega 040809a$, $\omega 040929b$, $\omega 041202a$, $\omega 041215b$, $\nu 061112a$, $\nu 061116b$, $\nu 070411a$, and $\nu 070412b$. All sessions were run in the Social Sciences Experimental Laboratory at Caltech, with Caltech students (undergraduate, graduate, and summer visiting students). Session $\omega 040929b$ had seventeen participants, sessions $\nu 061112a$ and $\nu 070411a$ had fourteen, session $\nu 061116b$ had sixteen participants, and session $\nu 070412b$ had eighteen participants; all other sessions had fifteen participants. In sessions of setup ω, the seven-minute time limit for Prize periods was never binding: participants always turned in their answer sheets early. Total earnings before sign-up reward fluctuated between \$0 and \$163. Mean payment amounted to \$31; the median payment was \$20.

Our results are taken from two sources. We have all the information on trades and prices that is collected

[13] There are exceptions, corresponding to cases where it is obvious if a certain object belongs or not to the optimal knapsack. For example, if one object weighs more than the sum of all remaining objects and its value is below the maximum value of the remaining objects, then this object clearly does not belong to the optimal knapsack. This cannot be extended to sets of more than one object - if a pair weighs more than the sum of the remainder and has a lower value than the maximal object in the remainder, this does not mean that both objects in the pair are not in the optimal knapsack. We make sure that none of the instances we consider falls in the class where the above described exception applies.

[14] If such an assignment of groups of knapsacks to subjects were exogenously given by a social planner, then this would be the team problem studied by Marschak and Radner [1972].

during market periods, plus we have the answer sheets of all participants for both treatments. These answer sheets are returned to the experimenter for no compensation. Participants are told at the beginning of the earning periods that they must give the answer sheets back to the experimenter, but their answers are completely irrelevant for payment, and participants understand this. Still, we are able to collect many answer sheets that are carefully filled out.[15] The number of answer sheets discarded because they are blank or incomprehensible is small, and similar across treatments for setup ω. For setup ν, the number of discarded answer sheets in the prize treatment is very small (approximately 2% of all answer sheets) while the number of discarded sheets (mainly because they are left blank) in the market treatment is large (approximately 25% of all answer sheets).

Result Zero: Instances are Hard. Before we move on to analyze the performance of markets vis-à-vis the prize treatment, we make the point that instances are *hard enough*, as defined in section 3.2.3. Proposition 8 says that it is only sensible to use Bayesian updating and thus the standard information aggregation paradigm, if instances are easy enough. If problems are easy enough, prices must immediately collapse to one or zero, depending on whether the object belongs or not to the optimal knapsack.

Direct evidence that problems are hard enough is the fact that only a fraction of participants solve them in both treatments. This is true for all instances (see table 3.3). Indirect evidence is the fact that prices never collapse to zero or one.

We can thus pursue our objective of understanding intellectual discovery outside the Bayesian learning framework. We present results from each setup in a separate section. The data for both setups demonstrate that markets do generate the solution to the instances of the knapsack problem, that trade is abundant and prices are informative. All results show a very strong correlation between the Sahni difficulty ranking of an instance and the number of participants that solve it.

3.4.1 Market Performance

Every instance that is solved under the prize treatment is also solved under the market treatment.

Table 3.3 shows the fraction of participants that find the correct solution for every instance of the knapsack problem. Setups ω and ν are displayed separately, and results are split according to the treatment. There is a big variance across instances in the number of correct solutions and instances that are more frequently solved under the market treatment are also more frequently solved under the prize treatment. Table 3.3 shows that all instances except instance VII are solved by a significant fraction of participants in both treatments. Percentages are taken with respect to the total number of participants, without adjusting for discarded answer sheets. The solution is always found in the market.

Figures 3.1 to 3.4 display the fraction of correct solutions and the number of choices of each object (correctly or incorrectly chosen) taken from participants' answer sheets. Results for setup ν are in figures

[15] Answer sheets are not always easy to comprehend, in which case we discard them. One example of behavior that leads us to think that answer sheets are carefully filled out is that as we pick up the answers, they often take two or three extra seconds to make sure they hand us the answer they desire. Another example is that answer sheets often have several iterated answers that are scratched, with an arrow or marker pointing to the "surviving" answer.

Table 3.3. Percentage of total number of participants that marked the correct answer to each instance on their answer sheets.

		Instance							
		I	II	III	IV	V	VI	VII	VIII
Fraction Correct (%) – Setup ω	Market	20.00	3.13	31.25	62.5	60.00	15.63	0.00	26.67
	Prize	18.75	10.00	20.00	33.33	34.38	3.33	0.00	12.50
Fraction Correct (%) – Setup ν	Market	32.14	14.71	11.76	26.47	32.14	5.88	3.57	21.43
	Prize	17.65	28.57	46.43	57.14	38.24	10.71	0.00	17.65

3.5 to 3.8. This is simply a graphical representation of the results in 3.3. It is also interesting to notice in these plots that whenever a large proportion of participants agrees about an object being in the optimal knapsack, this object truly is. Participants are not misled in the markets.

How does the market do what it does? It could be the case that participants in the market treatment just sat down and tried as hard as in the prize treatment to solve the instance, completely disregarding the market. This is not the case, as attested by the following two results:

3.4.1.1 Market Result One: Trade and Bidding are Abundant.

Table 3.4 lists offer and trade statistics. Per-person and per-period averages are displayed, as well as per-asset and per-period averages. Standard errors are in parentheses. Across all Market periods in all sessions, a large number of offers and trades are recorded.

Table 3.5 splits statistics across IN and OUT securities. "IN" securities correspond to items that are in the optimal knapsack; "OUT" securities correspond to items that are not in the optimal knapsack. Table 3.6 displays trading activity per instance. Each instance is solved under the Market treatment in two experimental sessions, which we refer to as Sessions One and Two. Since the number of items, and hence, securities, differs across instances, table 3.6 also reports per-security averages across both Sessions.

Differences in the number of trades and orders are evident across instances, and across securities (IN vs. OUT securities). We can only speculate about the reason of these differences. That is the matter of section 3.5. It also catches the eye that the number of bids and asks is much larger in setup ν. This may be driven by the use of the simpler trading interface of jMarkets.

3.4.1.2 Market Result Two: Prices are Noisy but Informative.

For all but one instance, the distribution of prices of IN securities *first-order stochastically dominates* the distribution of prices of OUT securities. Transaction prices reveal information about the optimal solution, but very noisily. The one instance where this dominance relation is not found, is instance VII, for which nobody finds the solution. First-order stochastic dominance in this context means that the probability that the price of an IN security is larger than the price of an OUT security is at least 0.5. Figure 3.9 shows the empirical distribution functions for prices of IN and OUT securities in every instance in setup ω, while figure

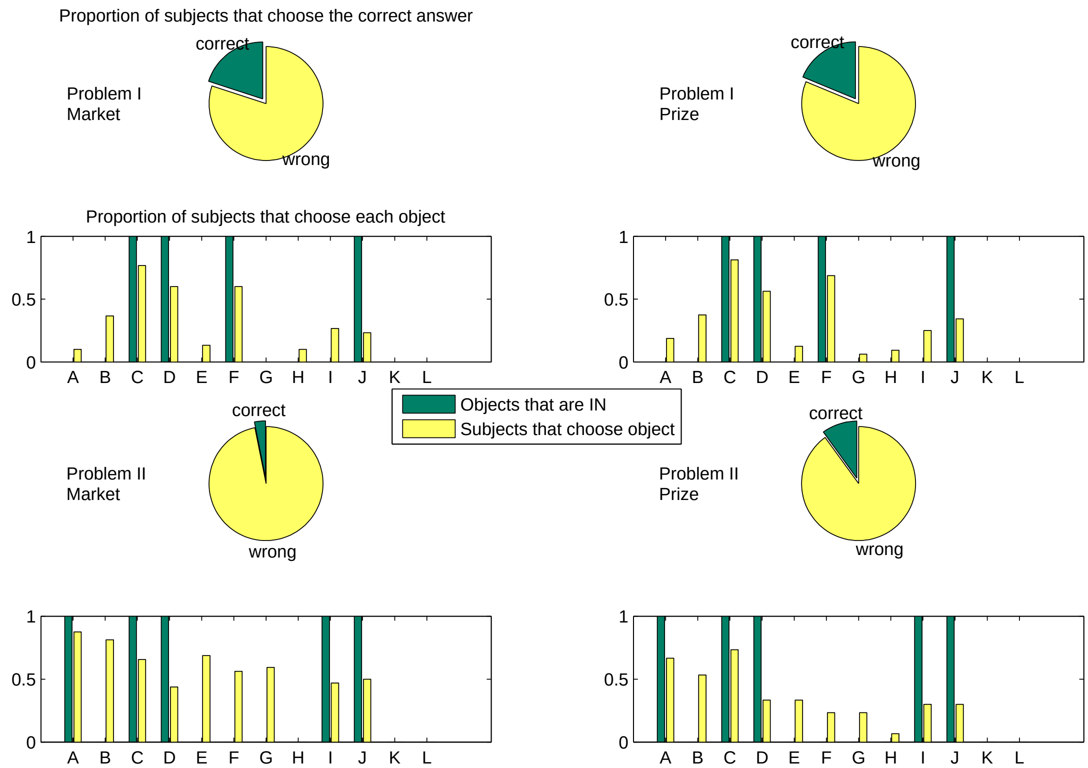


Figure 3.1. Answer-sheets data. Instances I and II.

Table 3.4. Summaries of trade and bidding activity. In setup ν all trades involve only one unit, so volume equals number of trades.

		Experiment							
		Setup ω				Setup ν			
		040809a	040929b	041202a	041215b	061112a	061116b	070411a	070412b
Average per person per period	Asks	6.97 (5.1)	5.38 (3.6)	7.68 (4.4)	5.3 (4)	25 (22)	22.7 (22)	36.9 (27.8)	22.1 (18)
	Bids	6.28 (3.7)	5.15 (2.9)	6.07 (4.2)	5.03 (3.7)	20 (28.8)	20.6 (26.9)	19.1 (9.4)	13.5 (10.9)
	Trades	5.43 (3.8)	4.19 (3.5)	5.63 (3.2)	4.07 (2.7)	10.6 (8.3)	8 (6.8)	13.3 (7.3)	9.1 (6.7)
	Volume	11.07 (8.5)	8.79 (6.9)	12.6 (7.8)	12.88 (9.2)	- -	- -	- -	- -
Average per asset per period	Asks	8.71 (4.1)	7.62 (3.6)	9.6 (4.8)	6.62 (3.9)	31.7 (25)	33 (18)	47 (33)	36.1 (27.2)
	Bids	7.85 (5.3)	7.29 (3.9)	6.07 (4.2)	6.29 (3.7)	25.5 (13.1)	30 (13.2)	24.4 (17.2)	22.1 (16)
	Trades	6.79 (3.8)	5.94 (4.0)	7.04 (3.2)	5.08 (3.8)	13.5 (7)	11.6 (7.6)	17 (10.5)	14.9 (10.6)
	Volume	13.83 (9.9)	12.46 (9.0)	15.75 (9.8)	16.1 (11.9)	- -	- -	- -	- -
Totals	Asks	418	366	461	318	1397	1451	2068	1591
	Bids	377	350	364	302	1123	1322	1072	975
	Trades	326	285	338	244	594	512	746	655
	Volume	664	598	756	773	-	-	-	-

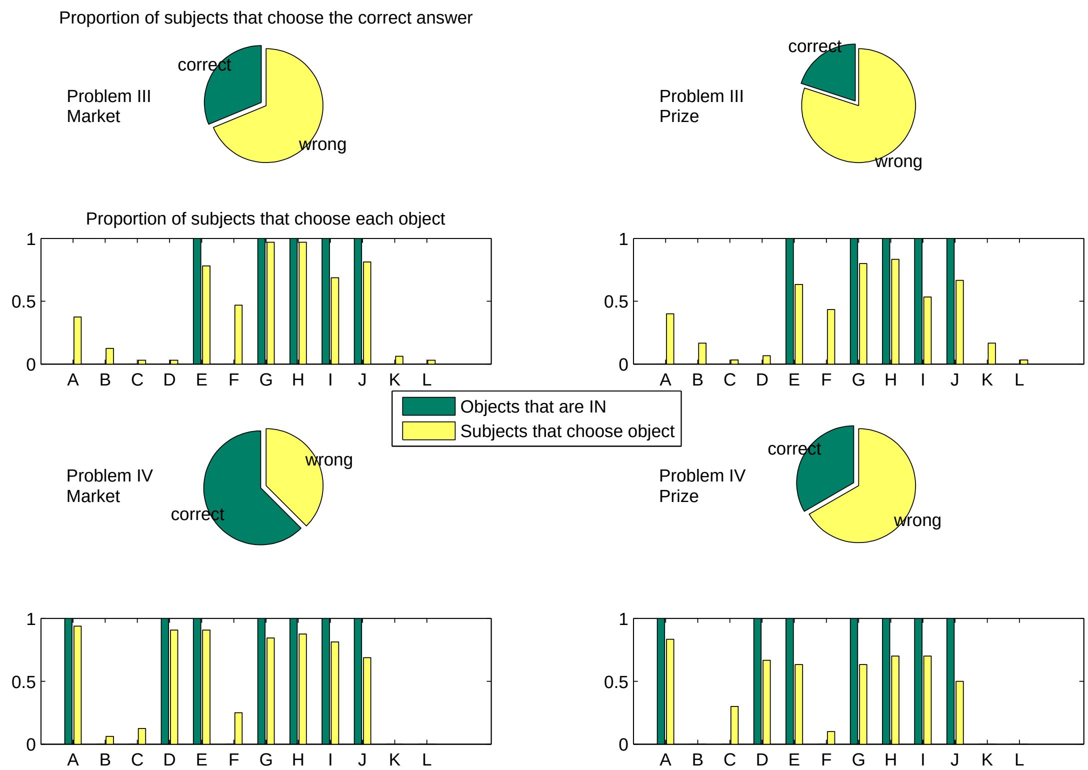


Figure 3.2. Answer-sheets data. Instances III and IV.

Table 3.5. Volume of trade and number of bids and asks, for IN and OUT securities.

		Experiment							
		Setup ω				Setup ν			
		040809a	040929b	041202a	041215b	061112a	061116b	070411a	070412b
Trade volume	IN	15.4	10.7	17.3	12.5	10.3	8.5	18	13.1
	OUT	14.9	17.7	17.12	24.9	16	15.8	16	17.5
Asks	IN	7.9	6.0	9.5	5.3	13.4	22.3	25.8	23.2
	OUT	10.7	11.7	11.2	10.1	45.7	46.2	63.1	54.9
Bids	IN	10.9	8.0	9.8	6.3	25	31.6	33	23.8
	OUT	6.8	7.9	7.1	7.7	18	27.8	17.8	19.8

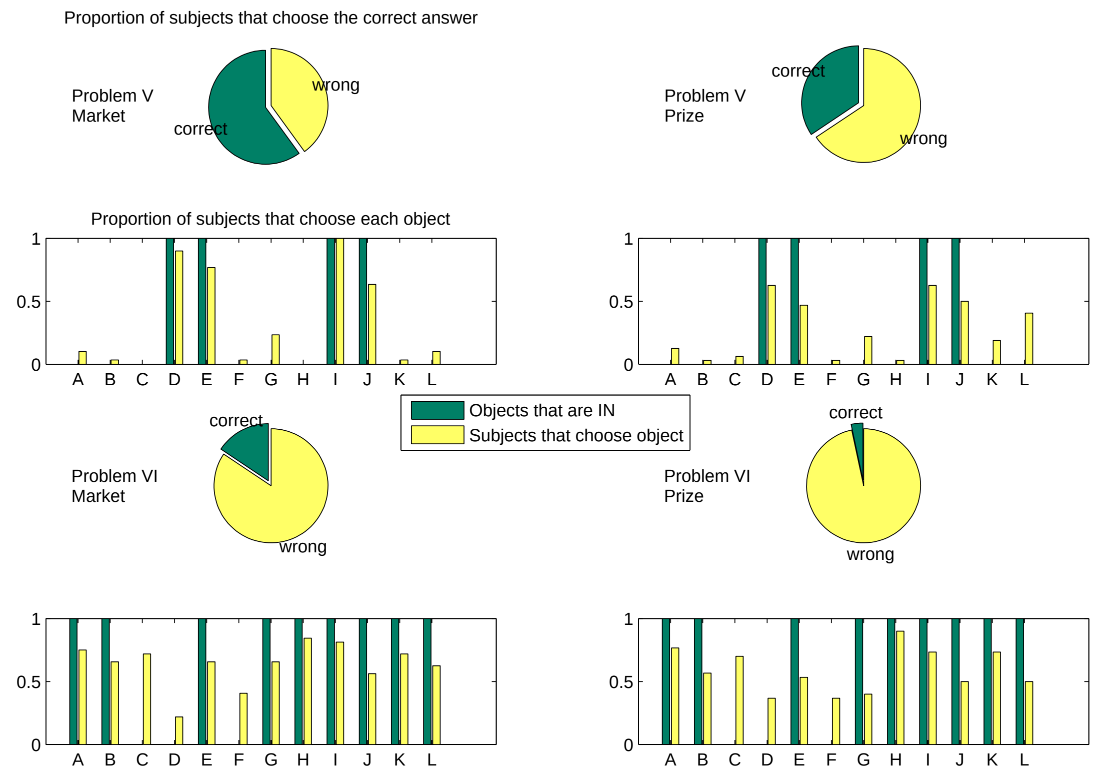


Figure 3.3. Answer-sheets data. Instances V and VI.

Table 3.6. Number of trades for each instance of the knapsack problem.

		Instance							
		I	II	III	IV	V	VI	VII	VIII
Setup ω	*Session One*	84	82	81	57	88	65	74	80
	Session Two	78	75	57	59	69	53	87	104
	avge. per asset	8.1	7.85	5.75	5.8	6.54	4.92	6.71	9.2
	n	10	10	12	10	12	12	12	10
Setup ν	*Session One*	135	138	140	142	171	87	114	174
	Session Two	122	240	140	183	193	92	197	234
	avge. per asset	12.8	18.9	11.7	16.3	15.2	7.5	12.9	20.4

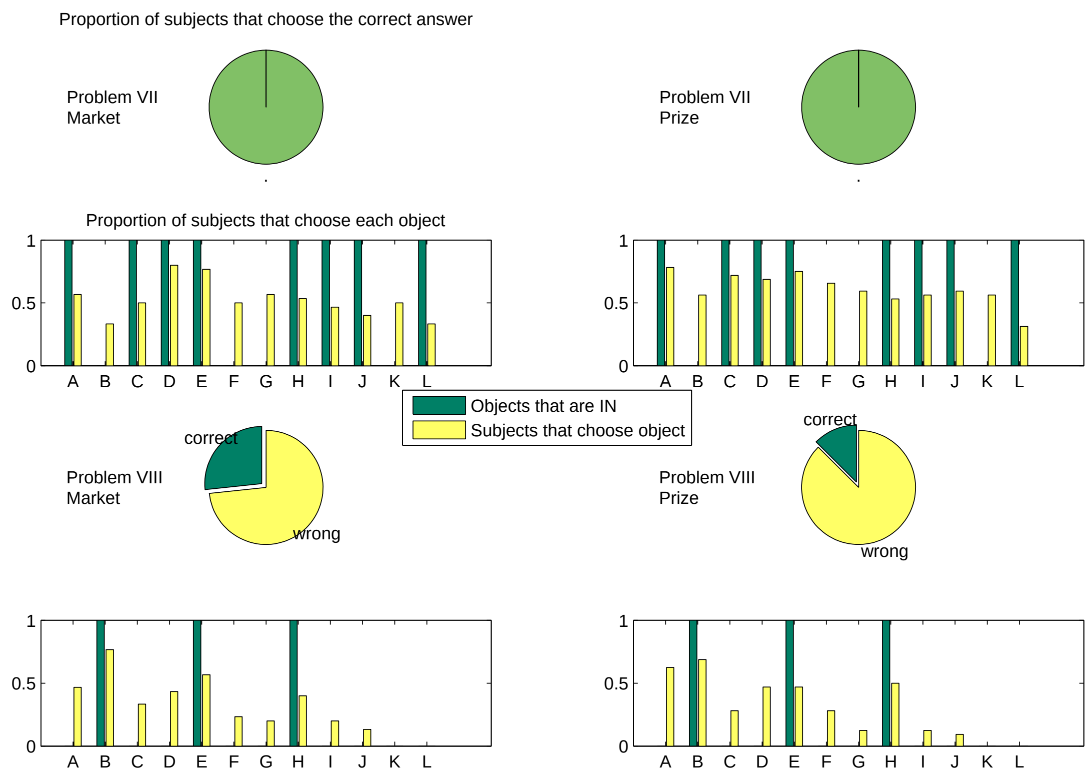


Figure 3.4. Answer-sheets data. Instances VII and VIII.

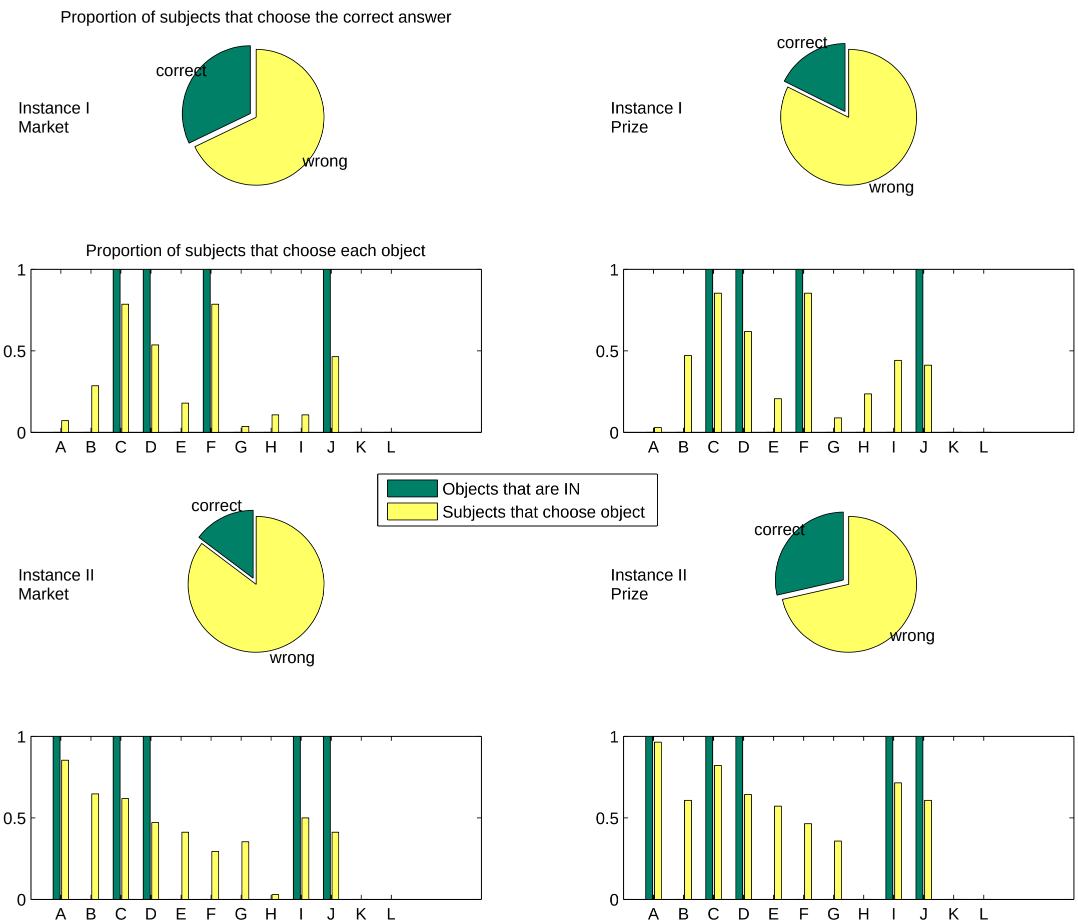


Figure 3.5. Answer-sheets data. Instances I and II. Setup ν.

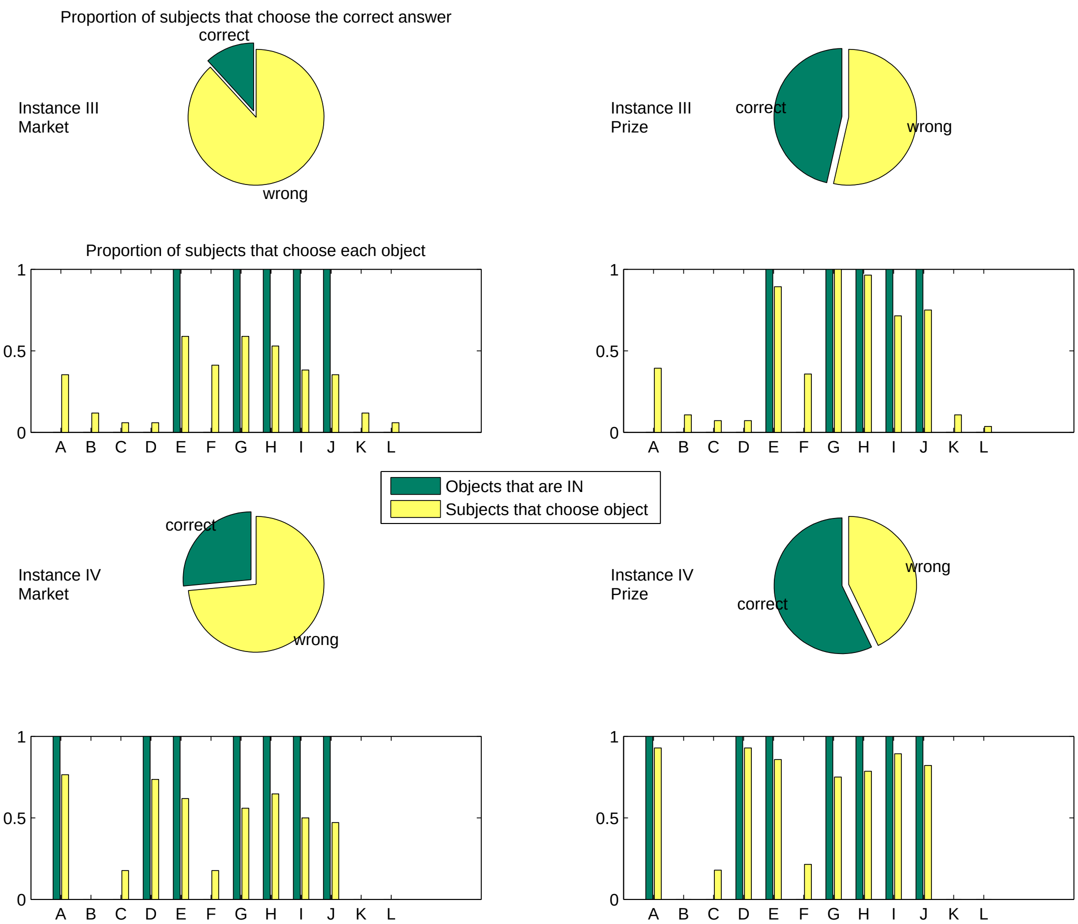


Figure 3.6. Answer-sheets data. Instances III and IV. Setup ν.

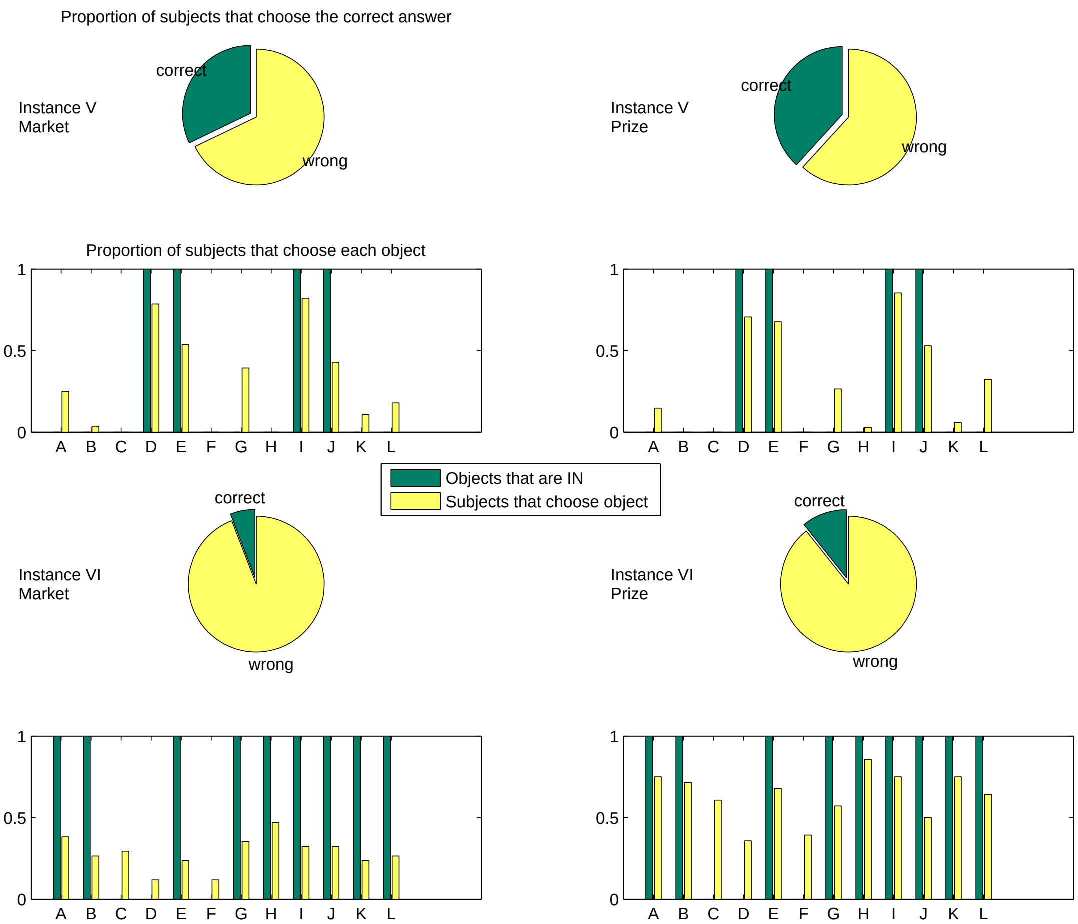


Figure 3.7. Answer-sheets data. Instances V and VI. Setup ν.

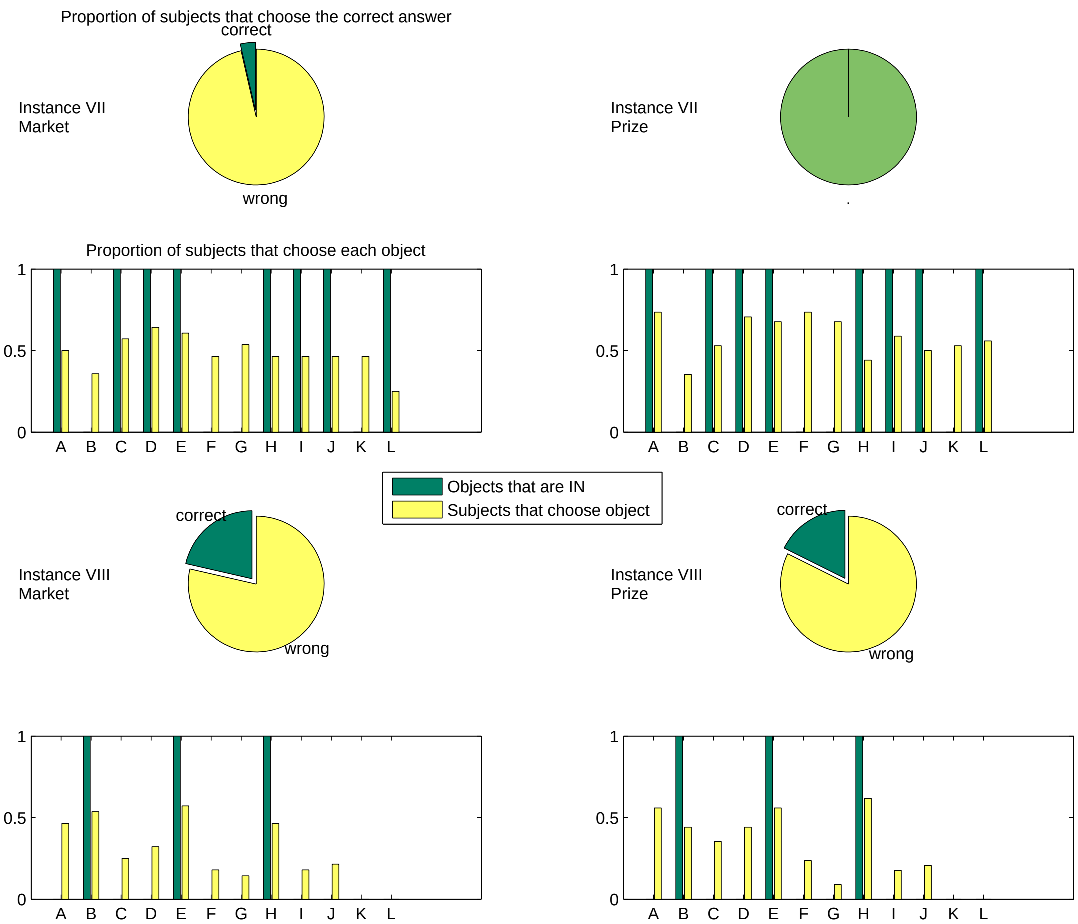


Figure 3.8. Answer-sheets data. Instances VII and VIII. Setup ν.

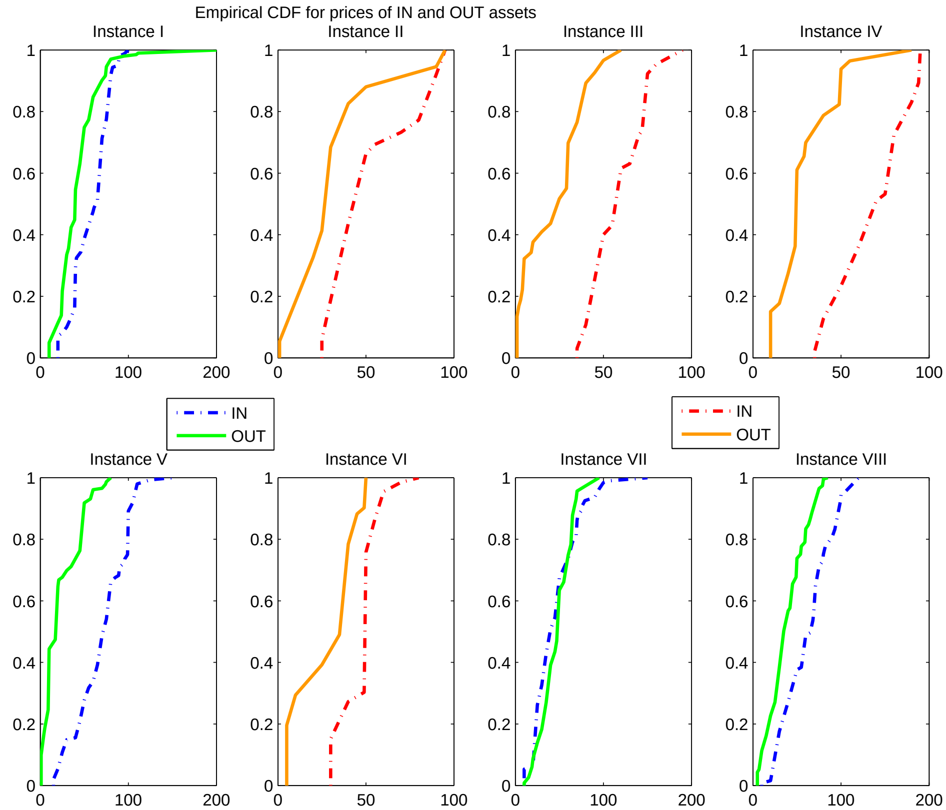


Figure 3.9. Empirical Cumulative Density Functions for prices of IN and OUT securities. Instances solved under the Market setup in experiment type a have a different color scheme than those solved under the Market setup in experiment type b. *Setup* ω

3.10 shows these distributions for setup ν.

There is clear evidence that participants do *interact* in the market. There is the potential for communication, and prices are an effective language since securities are differentiable based on their prices.

Finally, we report two more characteristics of our experimental markets which might carry some information but are harder to interpret.

- **Prices are low.** In setup ω, the median price is \$0.5, while the average price is \$0.48. Average and median prices of IN and OUT securities lie slightly above and below \$0.5, respectively. Both the median and the average become even lower when we add data from setup ν sessions.

- **Excess supply of OUT securities.** For every trade, we know whether it was the result of a buyer's submission of a market order (an order that takes an offer that is already waiting in the book and transforms it into a trade), or of a seller's submission of a market order. Among trades that are the result of a buyer's market order, a significantly higher number correspond to OUT securities than to IN securities. Also, most trades of OUT securities are the result of a buyer's market order. The reverse

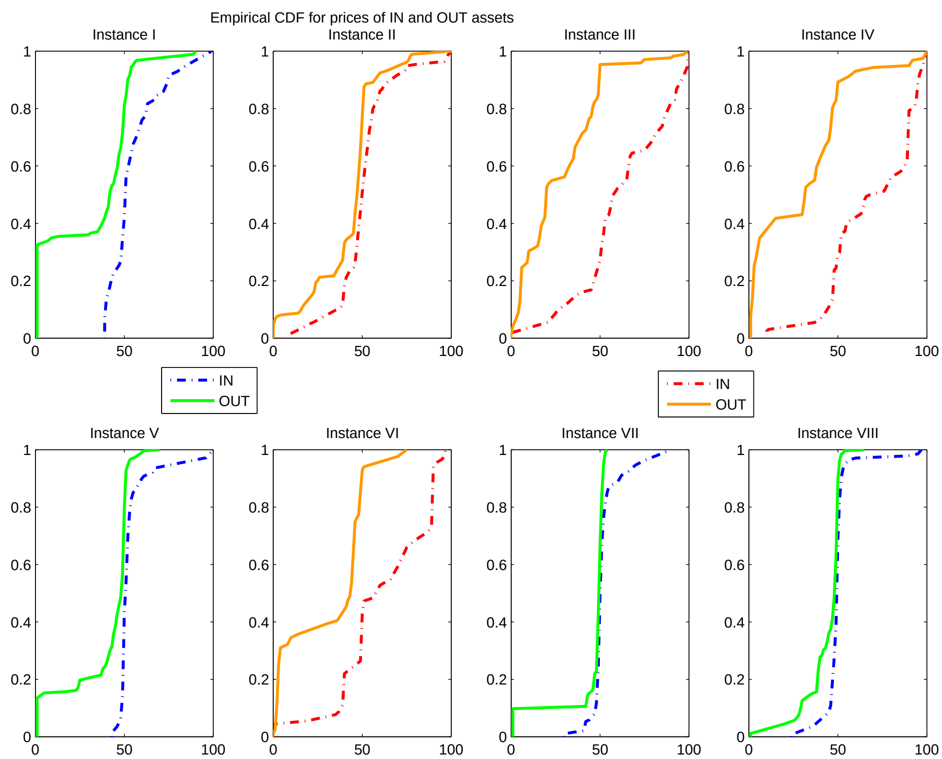


Figure 3.10. Empirical Cumulative Density Functions for prices of IN and OUT securities. Instances solved under the Market setup in experiment type a have a different color scheme than those solved under the Market setup in experiment type b. *Setup* ν.

is not true for IN securities.

3.4.2 Sahni-k is a Good Measure of Difficulty

If we consider only setup ω, in both treatments, Sahni-k difficulty is a good predictor of the number of correct answers for an instance. We propose and describe this measure in section 3.2.2. Here we summarize the ranking of instances that follows from using this measure:

$$IV \prec V \sim I \prec VIII \sim III \prec II \sim VI \prec VII,$$

where $\prec$ indicates that an instance is "easier," and $\sim$ indicates that two instances are not ranked, according to the Sahni-k difficulty measure.

From table 3.3 we get the following ordering of instances in the market (setup ω):

$$\tilde{I}_{IV} > \tilde{I}_{V} > \tilde{I}_{III} > \tilde{I}_{VIII} > \tilde{I}_{I} > \tilde{I}_{VI} > \tilde{I}_{II} > \tilde{I}_{VII},$$

where $\tilde{I}_s$ denotes the fraction of participants that solve instance s correctly. Instances in the prize treatment rank very similarly,

$$\tilde{I}_{V} > \tilde{I}_{IV} > \tilde{I}_{III} > \tilde{I}_{I} > \tilde{I}_{VIII} > \tilde{I}_{II} > \tilde{I}_{VI} > \tilde{I}_{VII}.$$

In both cases the ranking agrees with the Sahni-k ranking for all instances except instance I, which is "harder" in the experiment than predicted by our difficulty measure.

The clear correlation we just mentioned disappears in setup ν. It is only approximately true in this case. In setup However, we argue that this is mainly caused by the fact that participants in one of the two sessions of this setup outperformed participants in the other session for every instance. This heterogeneity between groups changes the relative ranking of instances that were solved in the market in one session versus those solved in the other session (analogously for instances solved in the prize). To see that this is indeed the case, notice that the ranking within a session and treatment (e.g., the instances solved under the market treatment in session $\nu 061112a$) displays the strong correlation with Sahni difficulty that was noted for setup ω.

3.5 Conclusion

We create in the laboratory a situation that emulates intellectual discovery with the property that it cannot be reasonably modeled as Bayesian learning. The results point out that this property may be relevant since in a market for discovery we find that: the discovery is made, trading activity is abundant, and prices are informative but noisy.

To exactly what extent these results are driven by the non-incrementality feature of the cognitive task we use to represent discovery must still be determined. We set the ground for the formulation of hypotheses that can be tested with further experiments.

In particular, the market treatment may deliver the results it does for several different reasons, which may actually interact. We mention three reasons that are worthy of and prone to further experimental study. First, it may be that participants find a code to coordinate on decentralized computation of the solution. Second, it may be that markets provide a "check," where participants try to corroborate whether the algorithm they are using for computation is right or wrong. Third, the markets do provide a way to reduce the risk that comes from attempting to guess the solution on a hunch (heuristic). This may make participants more willing to express their hunches, by means of which prices become informative. This reason may interact with one or both of the previous reasons. It can be studied further by changing the payoff in the market to incorporate the risk of being paid only if one is entirely correct (e.g., have securities pay dividends only if someone finds the solution, or pay per optimal solution held in the final portfolio, not per security).

We have given most thought to the pursuit of the first reason mentioned above. One check of the idea that participants manage to distribute computation is to verify if markets do better in a setup where coordination is easier, while the task is still computationally difficult. Another option is to move away from the cognitive task and instead set up an environment where the selective acquisition of information by different participants may benefit from coordination. In this purely informational framework, will coordination ensue even when it is very difficult to encode (as is the case in our experiment)?

Another contribution of this paper is to bridge a connection between certain notions of computer science and economic decisions where incentives matter. The description of non-incrementality of a problem in terms of solution algorithms, and the predictive power of the Sahni difficulty measure indicate that there is a bridge. In particular, it is our belief that the theory of intellectual discovery can gain much from the models of individual cognition that are inferred from the study of algorithms.

Appendix A

Prices and Trades in Dynamic Completeness Experiment

Average TpM taken over:	Experimental Session 051005 TpM	051005 $\frac{TpM}{\#subjects}$	051022 TpM	051022 $\frac{TpM}{\#subjects}$	051122 TpM	051122 $\frac{TpM}{\#subjects}$	051201 TpM	051201 $\frac{TpM}{\#subjects}$
Complete mkts.	9.6	0.8	14.2	0.65	28.4	0.9	21.6	0.7
Incomplete mkts. before announce	8	0.67	12.6	0.57	26.5	0.9	23	0.74
Incomplete mkts. after announce	5.8	0.48	8.8	0.4	19.3	0.6	16	0.52

Table A.1. Average number of trades per minute (TpM) and number of trades per minute, per participant.

Trades per subject in periods:	Experimental Session 051005 *Avge.*	051005 *Std. Dev.*	051022 *Avge.*	051022 *Std. Dev.*	051122 *Avge.*	051122 *Std. Dev.*	051201 *Avge.*	051201 *Std. Dev.*
Complete mkts.	28.7	14.2	23.2	13.6	33	26.2	25	19.8
Incomplete mkts. before announce	26.7	15.4	22.9	16.7	34.1	28.9	29.7	27.3
Incomplete mkts. after announce	19.3	12.7	15.9	11.6	24.8	19.4	20.7	18

Table A.2. Average and Standard deviation of the number of trades that a participant is involved in during an experimental session.

In this appendix we describe the details of the experimental sessions that we ran. We give statistics on trade and price behavior that are informative, but not central to the point we make in the paper.

The average supply of assets (market portfolio) differed slightly across sessions. In our experimental sessions, agents were of two possible types, depending on their initial holdings of assets. The number of agents of each type determines the market portfolio. Table 2.3 describes the initial holdings of each type as well as the ensuing market portfolio for each experimental session.

Tables A.1 and A.2 contain statistics on trading volume. Table A.1 contains per minute average trading volume for each experimental session. Because different experiments have different numbers of participants, the per minute values are divided by the number of participants and reported. The trading volume averages are taken over complete market periods, incomplete market periods *before* announcement, and incomplete market periods *after* announcement.

Table A.2 reports the average number of trades that each participant is involved in, as well as the standard deviation across participants of this number (notice that each trade involves two participants, hence the discrepancy between numbers in the two tables, since Table A.2 doubles the number of trades from Table A.1). Standard deviation is computed for the experimental totals over complete market periods, and incomplete market periods, both before and after announcement. Naturally, standard deviations in single periods are significantly larger than standard deviations of the totals, which are reported here.

The evolution of trading prices of tradeable assets over time is summarized in Figures A.1 to A.4.

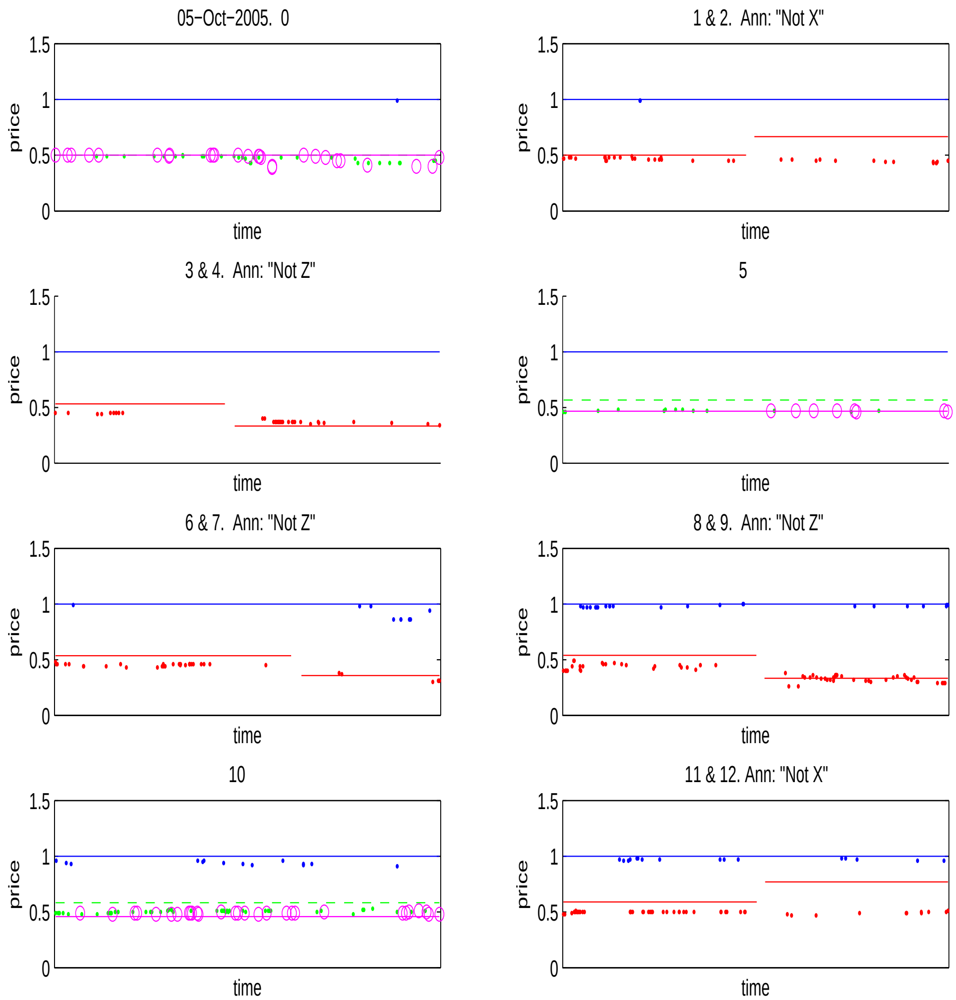


Figure A.1. Price evolution in time for every period. Straight lines indicate expected values of security dividends, given the contents of the urn and the announcement in each period. In complete market periods, circles and a solid line correspond to asset A, while dots and a dashed line correspond to asset B.

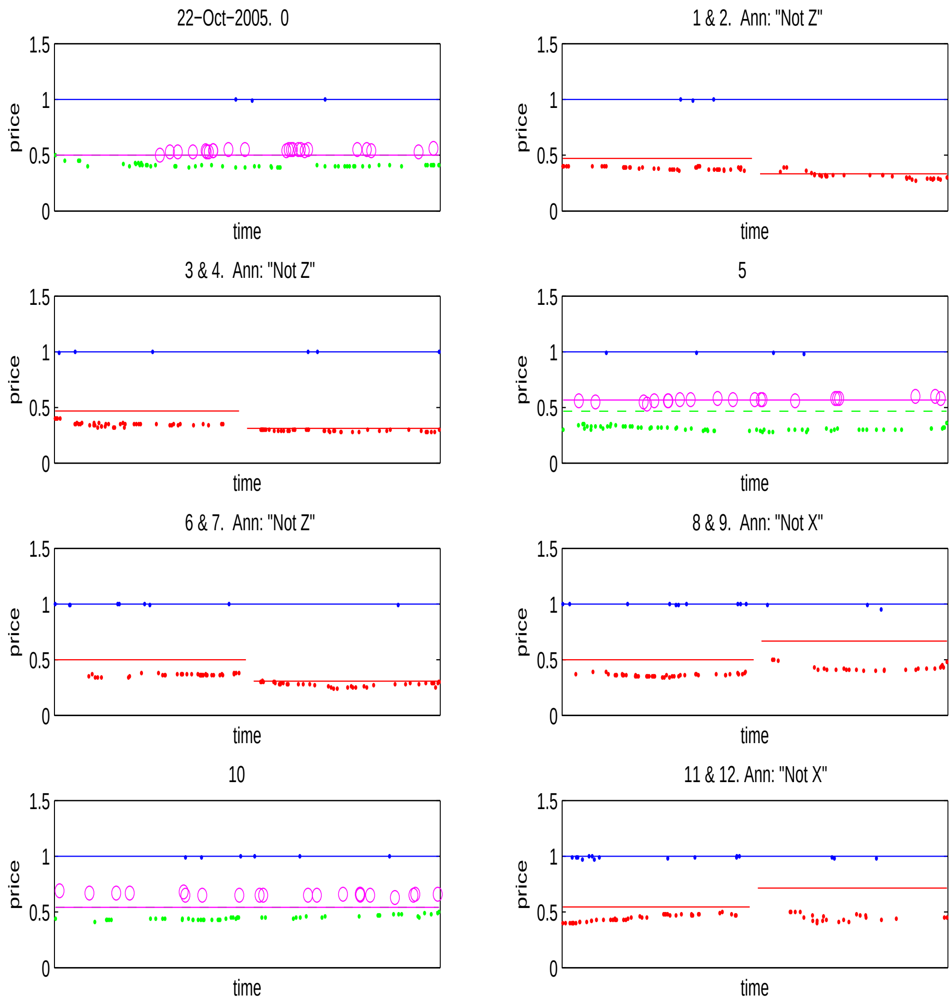


Figure A.2. Price evolution in time for every period. Straight lines indicate expected values of security dividends, given the contents of the urn and the announcement in each period. In complete market periods, circles and a solid line correspond to asset A, while dots and a dashed line correspond to asset B.

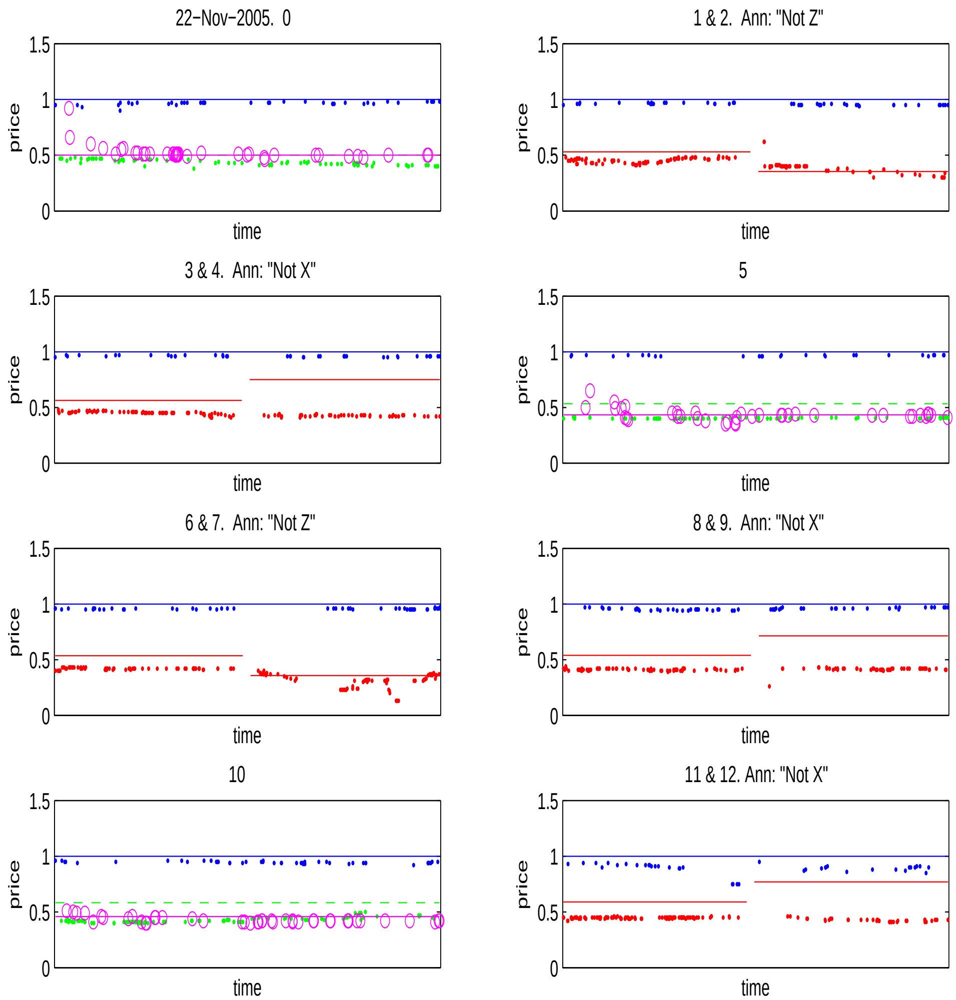


Figure A.3. Price evolution in time for every period. Straight lines indicate expected values of security dividends, given the contents of the urn and the announcement in each period. In complete market periods, circles and a solid line correspond to asset A, while dots and a dashed line correspond to asset B.

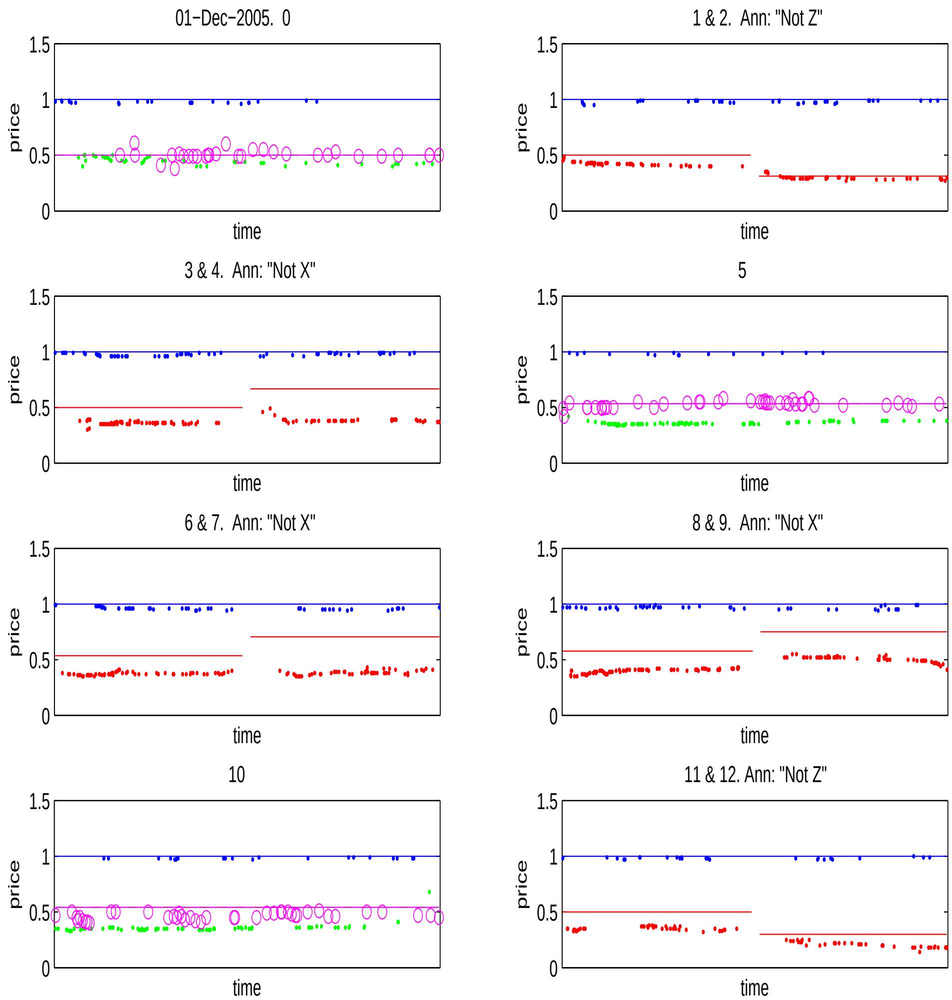


Figure A.4. Price evolution in time for every period. Straight lines indicate expected values of security dividends, given the contents of the urn and the announcement in each period. In complete market periods, circles and a solid line correspond to asset A, while dots and a dashed line correspond to asset B.

Bibliography

Enriqueta Aragones, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler. Fact-free learning. *The American Economic Review*, 95(5):1355–1368, 2005.

Kenneth Arrow. Methodological individualism and social knowledge. *The American Economic Review*, 84 (2):1–9, 1994.

Kenneth Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29(3): 155–173, 62.

Jonathan Berk. Necessary conditions for the capm. *Journal of Economic Theory*, 73:245–257, 1997.

Alan Blinder and John Morgan. Are two heads better than one? monetary policy by committee. *Journal of Money, Credit, and Banking*, 37(5), 2005.

Peter Bossaerts, Charles Plott, and William Zame. Prices and portfolio choices in financial markets: Theory and experiment. Mimeo, 2005.

Peter Bossaerts, Paolo Ghirardato, Serena Guarnaschelli, and William Zame. Prices and allocations in asset markets with heterogeneous attitudes towards ambiguity. MIMEO, 2007.

James Cox and Stephen Haynes. Barking up the right tree: Are small groups rational agents? *Experimental Economics*, 9:209–222, 2006.

G.B. Dantzig. Discrete variable extremum problems. *Operations Research*, 5:266–277, 1957.

Darrell Duffie and Chi-Fu Huang. Implementing arrow-debreu equilibria by continuous trading of few long-lived securities. *Econometrica*, 53(6):1337–1356, 1985.

Eugene Fama and Kenneth French. Data library, 2004. URL `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Nancy Gallini and Suzanne Scotchmer. Intellectual property: When is it the best incentive system? In Adam Jaffe, Joshua Lerner, and Scott Stern, editors, *Innovation Policy and the Economy*, volume 2, pages 51–78. MIT Press, 2002.

Michael Gibbons. Multivariate tests of financial models: A new approach. *Journal of Financial Economics*, 10:3–27, 1982.

Michael Gibbons, Stephen Ross, and Jay Shanken. A test of the efficiency of a given portfolio. *Econometrica*, 57(5):1121–1152, 1989.

Sanford Grossman. The existence of futures markets, noisy rational expectations, and informational externalities. *The Review of Economic Studies*, 44(3):431–449, 1977.

Sanford Grossman and Joseph Stiglitz. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408, 1980.

Boyan Jovanovic. Job matching and the theory of turnover. *The Journal of Political Economy*, 87(5): 972–990, 1979.

Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer-Verlag, 2004.

David Kreps. Multiperiod securities and the efficient allocation of risk: A comment on the black-scholes option pricing model. In J. McCall, editor, *The Economics of Uncertainty and Information*. University of Chicago Press, 1982.

Thomas Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

John Ledyard. Designing information markets for policy analysis. Presentation Slides, 2005.

Boris Maciejovsky and David Budescu. Is cooperation necessary? learning and knowledge transfer in cooperative groups and competitive auctions. MIMEO, 2005.

Jacob Marschak and Roy Radner. *Economic Theory of Teams*. Yale University Press, 1972.

S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, 1990.

Charles Plott and Shyam Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56(5):1085–1118, 1988.

S. Sahni. Approximate algorithms for the 0-1 knapsack problem. *Journal of the ACM*, 22:115–124, 1975.