Essays in Capital Markets

by

Dimitris Papanikolaou

B.A., Economics and Finance, University of Piraeus, 2000

Submitted to the Alfred P. Sloan School of Management in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

© Massachusetts Institute of Technology 2007. All rights reserved.

Author

Alfred P. Sloan School of Management

June 1, 2007

Certified by

Leonid Kogan

Associate Professor of Finance

Thesis Supervisor

Accepted by

Birger Wernerfelt

Professor of Management Science, Chair, Doctoral Program
Essays in Capital Markets

by

Dimitris Papanikolaou

Submitted to the Alfred P. Sloan School of Management
on June 1, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management

Abstract

In the first chapter, I provide evidence that investment-specific technological change is a source of systematic risk. In contrast to neutral productivity shocks, the economy needs to invest to realize the benefits of innovations in investment technology. A positive shock to investment technology is followed by a reallocation of resources from consumption to investment, leading to a negative price of risk. A portfolio of stocks that produce investment goods minus stocks that produce consumption goods (IMC) proxies for the shock and is a priced risk factor. The value of assets in place minus growth opportunities falls after positive shocks to investment technology, which suggests an explanation for the value puzzle. I formalize these insights in a dynamic general equilibrium model with two sectors of production. The model's implications are supported by the data. The IMC portfolio earns a negative premium, predicts investment and consumption in a manner consistent with the theory, and helps price the value cross section.

In the second chapter, based on joint work with Igor Makarov, we use heteroscedasticity of stock returns as an identification tool to isolate four robust factors in the U.S. industry returns. The first factor can be viewed as a proxy for economy-wide demand shocks. The second factor is a portfolio of stocks producing investment goods minus stocks producing consumption goods (IMC). The third factor differentiates between cyclical vs. non-cyclical stocks. Finally, the fourth factor is consistent with a proxy for shocks to input good prices. The extracted factors are shown to be important in explaining the cross-section of expected returns. Unlike the CAPM or the Fama and French three factor model, they successfully price the cross-section of 48 industry portfolios and do a good job at explaining the 25 Fama and French size and book-to-market portfolios. The fourth ("input") factor is found to be a robust predictor of the value-weighted market portfolio.

In the third chapter, based on joint work with Jiro Kondo, we propose a new foundation for the limits to arbitrage based on financial relationships between arbitrageurs and banks. Financially constrained arbitrageurs may choose to seek additional financing from banks who can understand their strategies. However, a hold-up problem arises because banks cannot commit to provide capital and have the financial technology to profit from the strategies
themselves. Wary of this, arbitrageurs will choose to stay constrained and limit their cor-
rection of mispricing unless banks have sufficient reputational capital. Using the framework
of stochastic repeated games, we show that this form of limited arbitrage arises when mis-
pricing is largest and becomes more substantial as the degree of competition between banks
intensifies and arbitrageur wealth increases.

Thesis Committee:
Leonid Kogan, Associate Professor of Finance, Chair
Stephen A. Ross, Franco Modigliani Professor of Financial Economics
Andrew W. Lo, Harris & Harris Group Professor; Director L.F.E.
Jun Pan, Mitsubishi Career Development, Associate Professor of Finance
Acknowledgments

I am grateful to my advisers Leonid Kogan, Andrew Lo, Jun Pan and Steve Ross for their guidance, to my colleagues Jiro E. Kondo, Igor Makarov, Isil Erel, Pierre Yared, Jack Bao and Alejandro Drexler for stimulating discussions, and most importantly, to Vasia Panousi for her constant support.

Additionally, I would like to thank George-Marios Angeletos, Darrell Duffie, Albert S. Kyle, Jon Lewellen, Roberto Rigobon, Tano Santos and Jiang Wang for helpful comments and discussions.
Contents

1 Investment-Specific Technological Change and Asset Prices ........................................ 13
  1.1 Introduction ................................................................................................................. 13
  1.2 Relation to the existing literature .................................................................................. 15
  1.3 General equilibrium model ......................................................................................... 17
      1.3.1 Information .......................................................................................................... 17
      1.3.2 Firms and technology ......................................................................................... 17
      1.3.3 Households .......................................................................................................... 19
      1.3.4 Competitive equilibrium ..................................................................................... 20
      1.3.5 Asset prices ......................................................................................................... 23
  1.4 Empirical evidence ........................................................................................................ 26
      1.4.1 Investment minus consumption portfolio ......................................................... 27
      1.4.2 Cross-sectional tests ......................................................................................... 29
      1.4.3 Does IMC proxy for investment-specific technological change? ..................... 32
  1.5 Conclusion ...................................................................................................................... 37
  1.6 Tables and figures ......................................................................................................... 38

2 Sources of systematic risk .................................................................................................. 55
  2.1 Introduction .................................................................................................................... 55
  2.2 Related literature .......................................................................................................... 58
  2.3 Econometric framework ............................................................................................... 59
      2.3.1 Model ................................................................................................................... 59
      2.3.2 Estimation ............................................................................................................ 60
      2.3.3 Empirical implementation ................................................................................... 62
  2.4 Empirical results ............................................................................................................ 62
      2.4.1 The number of factors ......................................................................................... 62
      2.4.2 Interpretation ....................................................................................................... 63
      2.4.3 Robustness .......................................................................................................... 65
      2.4.4 Asset pricing tests .............................................................................................. 65
      2.4.5 Predictive regressions .......................................................................................... 68
  2.5 Link to Macroeconomic Variables .................................................................................. 70
  2.6 Beta dynamics ................................................................................................................. 71
      2.6.1 Empirical Test ...................................................................................................... 72
  2.7 Discussion ....................................................................................................................... 74
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>IMC</td>
<td>41</td>
</tr>
<tr>
<td>1-2</td>
<td>CAPM pricing errors vs covariance with IMC</td>
<td>48</td>
</tr>
<tr>
<td>1-3</td>
<td>Cumulative response on IMC: Investment</td>
<td>50</td>
</tr>
<tr>
<td>1-4</td>
<td>Cumulative response on IMC: Consumption and output</td>
<td>51</td>
</tr>
<tr>
<td>2-1</td>
<td>Daily stock returns for the NASDAQ and NYSE indices</td>
<td>90</td>
</tr>
<tr>
<td>2-2</td>
<td>Pricing errors. 48 Fama and French industry portfolios</td>
<td>91</td>
</tr>
<tr>
<td>2-3</td>
<td>Pricing errors. 25 Fama and French B/M and Size sorted portfolios</td>
<td>92</td>
</tr>
<tr>
<td>2-4</td>
<td>Betas comovement</td>
<td>93</td>
</tr>
<tr>
<td>2-5</td>
<td>Conditional Variance of $f_i$</td>
<td>94</td>
</tr>
<tr>
<td>2-6</td>
<td>Model implied betas</td>
<td>95</td>
</tr>
<tr>
<td>3-1</td>
<td>An arbitrage opportunity cycle</td>
<td>102</td>
</tr>
<tr>
<td>3-2</td>
<td>Illustration of Proposition 6</td>
<td>107</td>
</tr>
<tr>
<td>3-3</td>
<td>Equilibrium price spreads</td>
<td>110</td>
</tr>
<tr>
<td>3-4</td>
<td>Relationship thresholds</td>
<td>115</td>
</tr>
<tr>
<td>A-1</td>
<td>Solution of the baseline model</td>
<td>136</td>
</tr>
<tr>
<td>A-2</td>
<td>Solution of model with 2 capital stocks</td>
<td>137</td>
</tr>
</tbody>
</table>
## List of Tables

1.1 IMC Portfolio: Composition ............................................. 38
1.2 IMC Portfolio: Correlations ........................................... 38
1.3 Summary Statistics ..................................................... 39
1.4 Time-Series Variation .................................................. 40
1.5 Summary Statistics: 10 portfolios sorted on IMC beta .............. 42
1.6 Cross-Sectional Tests: 24 portfolios sorted on Industry and IMC covariance ......................................................... 43
1.7 Cross-Sectional Tests: 25 portfolios sorted on IMC and MKT beta ................................................................. 44
1.8 Cross-Sectional Tests: 25 portfolios sorted on ME and IMC covariance ............................................................ 45
1.9 Cross-Sectional Tests: 25 portfolios sorted on BM and ME ........ 46
1.10 Cross-Sectional Tests: 25 portfolios sorted on BM and IMC covariance ............................................................. 47
1.11 IMC and Expected Returns - Entire Cross-section of stocks .... 49
1.12 Real rates and IMC(X) .................................................... 52
1.13 IMC and Investment: Firm Level Evidence .......................... 53
2.1 The number of factors in 29 Fama and French industry portfolios . 76
2.2 Specification check ....................................................... 76
2.3 Factor loadings and correlation of 29 Fama-French industries for \( f_t \) .............................................................. 77
2.4 Correlation of \( f_t \) with other returns .................................. 78
2.5 Factor Interpretation ....................................................... 78
2.6 Robustness check .......................................................... 79
2.7 Correlation between \( f_t \) and \( f_t^f \) ........................................... 79
2.8 Factor loadings and correlation of 29 Fama-French industries for \( f_t^f \) and IMC .......................................................... 80
2.9 Summary statistics for \( f_t^f \) and IMC ..................................... 81
2.10 Alphas and betas for \( f_t^f \) and IMC ...................................... 82
2.11 Time-series Tests .......................................................... 83
2.12 Factors’ Cross-Autocorrelations ....................................... 83
2.13 Regressions of monthly market excess returns on lagged conditioning variables and \( f_t^f \) .......................................................... 84
2.14 Regressions of monthly market excess returns on lagged conditioning variables, \( f_t^4 \), and industry portfolios ........................................ 85
2.14 (Continued): Regressions of monthly market excess returns on lagged conditioning variables, \( f_t^4 \), and industry portfolios ........................................ 86
2.15 Regressions of monthly market excess returns on lagged conditioning variables, $f_t^*$, and commodity returns ............................................ 87
2.16 Regressions of market excess returns on lagged $f_t^*$ ......................... 88
2.17 Out-of-sample forecasting power of $f_t^*$ ........................................ 88
2.18 GARCH(1,1) estimates of $f_t$ ....................................................... 89

A.1 Unconditional Moments .......................................................... 135
Chapter 1

Investment-Specific Technological Change and Asset Prices

Abstract
In this chapter I provide evidence that investment-specific technological change is a source of systematic risk. In contrast to neutral productivity shocks, the economy needs to invest to realize the benefits of innovations in investment technology. A positive shock to investment technology is followed by a reallocation of resources from consumption to investment, leading to a negative price of risk. A portfolio of stocks that produce investment goods minus stocks that produce consumption goods (IMC) proxies for the shock and is a priced risk factor. The value of assets in place minus growth opportunities falls after positive shocks to investment technology, which suggests an explanation for the value puzzle. I formalize these insights in a dynamic general equilibrium model with two sectors of production. The model's implications are supported by the data. The IMC portfolio earns a negative premium, predicts investment and consumption in a manner consistent with the theory, and helps price the value cross section.

1.1 Introduction
The second half of the last century has seen remarkable technological innovations, the majority of which have taken place in equipment and software. Technological innovations affect output only to the extent that they are implemented through the formation of new capital stock. Since firms need to invest in order to benefit from advances in investment technology, these innovations do not necessarily benefit all firms equally. In this chapter, I argue that investment-specific technological change is a source of systematic risk that is responsible for some of the cross-sectional variation in risk premia, both between different sectors in the
economy and between value and growth firms.

I propose a dynamic general equilibrium model that links investment-specific technological change to asset prices. In contrast to the standard one-sector model, shocks to investment technology (I-shocks) do not affect the production of the consumption good directly. Instead, they alter the real investment opportunity set in the economy by lowering the cost of new capital goods. Since the old capital stock is unaffected, the economy must invest to realize the benefits. As the economy trades off current versus future consumption, there is a reallocation of resources from the production of consumption goods to investment goods. Because the economy is willing to give up consumption today, the marginal value of a dollar must be high in these states of the world. Therefore, stocks are more expensive if they pay off in states when real investment opportunities are good, or equivalently the investment-specific shock has a negative premium. The types of firms that are likely to do well in these states are firms that produce capital goods and firms with a lot of growth opportunities.

My model provides novel empirical implications about the cross-section of stock returns. Since investment-specific technological change is not directly observable, direct empirical tests are difficult to implement. However, one of the advantages of the model is that it provides restrictions which help identify investment-specific technological change using the cross-section of stock prices. In particular, the model features two sectors of production, one producing the consumption good and one producing the investment good. In this context, investment-specific technological change benefits firms that produce capital goods relative to firms that produce consumption goods. As a result, the investment shock is spanned by a portfolio of stocks producing investment goods minus stocks producing consumption goods (IMC). I construct an empirical equivalent of this portfolio and use it as a proxy for investment technology shocks.

The first implication of the model is that the investment-specific shock carries a negative premium. This implies that firms that are positively correlated with investment-specific shocks should have lower average returns. I show that sorting individual firms into portfolios on covariances with the IMC portfolio generates a spread in average returns that is not explained by the (C)CAPM or the Fama-French three factor model. On the other hand, a two factor model which includes the IMC portfolio in the (C)CAPM successfully prices the spread. In addition, I employ as test assets portfolios of firms sorted by firm characteristics and covariances with IMC and estimate the parameters of the pricing kernel. Finally, I repeat the procedure using the entire cross-section of stock returns, following Fama and French (1992). Overall, I find the estimates of the risk premium to be negative and very similar regardless of the test assets used.

The second implication is that the value of assets in place minus the value of growth opportunities has a negative correlation with the I-shock. This is important because it offers a novel explanation for the value puzzle. Specifically, a positive I-shock lowers the cost of new investment, which causes the value of future growth opportunities to increase and the value of assets in place to fall. As a result, growth stocks have lower expected returns because they do well in states where real investment opportunities are good and the marginal value of wealth is high. I find that including the IMC portfolio in the (C)CAPM dramatically
improves the ability of the model to price the cross-section of stocks sorted by book to market.

The model identifies the IMC portfolio as a proxy for investment-specific technological change. I examine this restriction more carefully in the following ways. I first show that positive returns on the IMC portfolio are followed by an increase in the quantity of investment and a fall in the quality-adjusted relative price of new equipment. This indicates that the IMC portfolio at least partially reflects a shock to the supply of investment. Next, I consider the model's predictions about consumption, namely that consumption falls in the short run but increases in the long run. Indeed, I find that the discretionary component of consumption falls following positive returns on IMC. However, the power to detect long-run responses of consumption is very low. Therefore, I examine the correlation of IMC with the real interest rate, which in the model is an increasing function of long-run consumption growth. I find that positive returns on IMC are correlated with changes in the long-term real interest rate, but not with changes in the short-term rate. Since monetary policy primarily affects the short-term rate, the correlation between returns on the IMC portfolio and real rates probably stems from its relation to long-term consumption growth, rather than a potential link with monetary policy.

The rest of the paper is organized as follows. Section 2 describes the related literature. Section 3 presents a simple general equilibrium model where the capital stock in the investment sector is fixed. Section 4 presents the empirical results. Section 5 concludes. The Appendix extends the basic model by allowing the economy to increase the capital stock in the investment sector. All technical details are relegated to the Appendix.

1.2 Relation to the existing literature

My work is motivated by two separate empirical observations, one in macroeconomics and one in finance. First, the macroeconomic literature has documented a negative correlation between the price of new equipment and new equipment investment. Greenwood, Hercowitz and Krussell (1997, 2000) interpret this finding as evidence for investment-specific technological change and show that it has the ability to explain both the long-run behavior of output as well as its short-run fluctuations. They calibrate an RBC model with investment-specific technological change, using the price of new equipment as a proxy for the realizations of the investment-specific shock. They show that this shock can account for a large fraction of both short-run and long-run output variability, in magnitudes of 30 to 60%. Fisher (2006) treats the shock as unobservable and uses a similar model to derive long-run identifying restrictions on a VAR, with the identified investment technology shock explaining up to 62% of output fluctuations over the business cycle. Second, in the finance literature, Makarov and Papanikolaou (2006) find that returns of industries producing final goods versus investment goods have different statistical properties. Specifically, they use an approach based on identification through heteroskedasticity to identify the latent factors affecting stock returns. One of the factors they recover is highly correlated with the investment minus consumption portfolio and HML. I provide a general equilibrium model that links these two facts by show-
ing that the first implies the second, and suggests an additional proxy for investment-specific shocks.

Jermann (1998) and Tallarini (2000) are early examples of work that explores the asset pricing implications of general equilibrium models. This literature builds on the neoclassical RBC model and focuses on aggregate quantities and prices. In this environment they find that the equity premium and risk-free rate puzzles are exacerbated because high risk aversion implies endogenously smooth consumption. The production economy model has been extended to allow for cross-sectional heterogeneity in firms, with the explicit purpose of linking firm characteristics such as book to market and size to expected returns. Zhang (2005) builds an industry equilibrium model and Gomes, Kogan and Zhang (2003), Gourio (2005), and Gala (2006) build general equilibrium models where investment frictions and idiosyncratic shocks result in ex-post heterogeneity in firms. My work is most closely related to Gomes, Kogan and Yogo (2006) who focus on ex-ante firm heterogeneity, that is heterogeneity arising because of differences in the type of firm output rather than differences in productivity or accumulated capital. They build a general equilibrium model where differences in the durability of a firm’s output lead to differences in expected returns, whereas I focus on the difference between capital good and final good producers. My paper is also close to theirs in terms of empirical methodology, since they also use the Input-Output tables from the Bureau of Economic Analysis to classify firms based on the type of output they produce. However, the general equilibrium models above have a single aggregate shock and therefore imply a one factor structure for the cross-section of stock returns. As a result, any difference in expected returns must be due to differences in market betas, since the conditional CAPM holds exactly. Even though true betas are unobservable, Lewellen and Nagel (2004) argue that they could not covary enough with the market premium to justify the observed premia. In addition, models with one shock cannot generate the pattern documented by Fama and French (1993), where value and growth stocks move together independent of the market portfolio. My model enriches the production technology of a standard general equilibrium model by differentiating between types of technological shocks.

The production technology in my model is different from the models above, but it has been extensively used in the macroeconomic literature. Investment-specific shocks were first considered by Solow (1960) in his growth model with vintage capital. Rebelo (1991) uses a two-sector AK model to study endogenous growth. Boldrin, Christiano and Fisher (2001) consider a similar model with two sectors of production, habit preferences and investment-specific shocks. They calibrate their model to match the equity premium, but their main focus is on improving on the quantity dynamics. In contrast, my interest is on deriving implications about the cross-section of stock returns. In the finance literature, Panageas and Yu (2006) build a general equilibrium model with different vintages of capital, but they focus on the comovement between asset returns and consumption over the long-run. In addition, this chapter is related to work that explores the effects of technological innovation and asset prices (Jovanovic (2001), DeMarzo, Kaniel and Kremer (2006), Pastor and Veronesi (2006)).

My paper is also related to recent work that explores the effect of the duration of a firm’s cashflows on expected returns, for example Campbell and Vuolteenaho (2004), Lettau and
Wachter (2006), Santos and Veronesi (2006b) and Lustig and Van Nieuwerburgh (2006). These models are based on the observation that growth stocks are higher duration assets than value stocks, and thus may be more sensitive to changes in the financial investment opportunity set in the spirit of Merton's ICAPM (1973). On the other hand, Bansal, Dittmar and Lundblad (2006) and Bansal, Dittmar and Kiku (2007) explore the fact that cashflows of growth and value firms have differential properties to explain the value premium. Because all these models are based on an exchange economy, changes in the financial investment opportunity set are either exogenously specified or arise through preferences. My work complements the papers above by considering time-varying real investment opportunities in a model with production.

Finally, my paper is related to the growing literature that explores the ability of the consumption-based model to explain the cross-section of expected returns. This literature focuses on measurement issues (Ait-Sahalia, Parker and Yogo (2004), Jagannathan and Wang (2005)), long horizons (Bansal, Dittmar and Lundblatt (2005), Parker and Julliard (2005), Malloy, Moskowitz and Vissing-Jorgenson (2006)), conditional versions of the CCAPM (Lettau and Ludvigson (2001), Santos and Veronesi (2006a)) or multiple good economies (Lustig and Van Nieuwerburgh (2005), Pakos (2004), Piazzesi, Schneider and Tuzel (2006), and Yogo (2006)). Similarly, my paper is also related to the literature that explores the empirical implications of production-based models, namely Cochrane (1991, 1996), Li, Vassalou and Xing (2006) and Belo (2006).

1.3 General equilibrium model

I build a general equilibrium model to formalize the idea that investment-specific shocks create a reallocation of resources between the consumption and the investment sector. The two-sector specification I consider is adapted from the model of Rebelo (1991) who studies endogenous growth. I first present a simplified version where the capital stock in the investment sector is fixed. In the Appendix, I solve the general model with two capital stocks and show that the main insights are robust.

1.3.1 Information

The information structure obeys standard technical assumptions. Specifically, there exists a complete \((\Omega, \mathcal{F}, \mathcal{P})\) probability space supporting the Brownian motion \(Z_t = (Z_t^X, Z_t^Y)\). \(\mathcal{P}\) is the corresponding Wiener measure, and \(\mathcal{F}\) is a right-continuous increasing filtration generated by \(Z\).

1.3.2 Firms and technology

Production in the economy takes place in two separate sectors, one producing the consumption good (numeraire) and one producing the investment good.
Consumption sector

The consumption goods sector (C-sector) produces the consumption good, \( C \), with two factors of production, sector specific capital \( K_C \) and labor \( L_C \)

\[
C_t \leq X_t K_C^\beta_c L_C^{-\beta_c},
\]

where \( \beta_c \) is the elasticity of output with respect to capital. Output in the C-sector is subject to a disembodied productivity shock \( X \) that evolves according to

\[
dX_t = \mu_X X_t dt + \sigma_X X_t dZ_t^X.
\]

The \( X \) shock can be interpreted as a neutral productivity shock (N-shock) that increases the productivity of all capital in the consumption sector. This is the standard shock in existing one-sector general equilibrium models. The capital allocated to the C-sector depreciates at a rate \( \delta \), while the investment in consumption-specific capital is denoted by \( I_C \). Thus, capital in the final goods sector evolves according to

\[
dK_{C,t} = I_{C,t} dt - \delta K_{C,t} dt.
\]

Investment in the C-sector is subject to adjustment costs. If the firm has capital \( K_C \) and wants to increase its capital by \( I_C \), it consumes \( c(-K_C) K_C \) total units of the investment good, where \( c(\cdot) \) is an increasing and (weakly) convex function. The value of a representative firm in the consumption sector equals

\[
S_{C,t} = E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( X_s K_C^{\beta_c} L_C^{-\beta_c} - w_s L_C S - \lambda_s c \left( \frac{I_{C,s}}{K_{C,s}} \right) K_{C,s} \right) ds,
\]

where \( w \) is the relative price of labor and \( \lambda \) is the relative price of the investment good, or equivalently the cost of new capital.

Investment sector

The investment goods sector (I-sector) produces the investment good using sector specific capital \( K_I \) and labor \( L_I \). In the simplified model, the capital stock in the investment sector is fixed. One can therefore think of the investment sector as using land and labor to produce the investment good. The output of the I-sector can be used to increase the capital stock in the C-sector

\[
c \left( \frac{I_{C,t}}{K_{C,t}} \right) K_{C,t} \leq \alpha Y_t K_I^{\beta_I} L_I^{-\beta_I}.
\]

The shock \( Y \), which represents the investment shock, affects the productivity of the investment sector. A positive shock to \( Y \) increases the productivity of the investment sector, which implies that the economy can produce the same amount of new investment \( (I_C) \) using fewer resources \( (L_I) \). Therefore, a positive investment shock will imply a fall in the cost of producing new capital, whereas the old capital stock will be unaffected. The elasticity of
output with respect to labor in the investment sector equals $1 - \beta_I$, and $\alpha$ is a parameter controlling the relative size of the two sectors. The investment shock follows

$$dY_t = \mu Y_t dt + \sigma Y_t dZ_t.$$

(1.6)

Firms in the investment sector represent claims on the land ($K_I$) used to produce investment goods. The value of a representative firm in the investment sector equals

$$S_{I,t} = \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( \lambda_s \alpha Y_s K_I^{\beta_I} L_{I,s}^{1-\beta_I} - w_s L_{I,s} \right) ds,$$

(1.7)

where $w$ is the wage and $\lambda$ is the relative price of the investment good in terms of the consumption good.

Finally, define the investment rate, $i_{C,s} = \frac{I_{C,s}}{K_{C,s}}$.

**Investment frictions**

Equation (1.5) implies that there is an upper bound on the investment rate in the consumption sector, because aggregate investment cannot exceed the output of the I-sector. This upper bound depends on labor allocated in the I-sector and the investment-specific shock $Y$. Hence the supply of capital is not perfectly elastic, and as a result future growth opportunities have positive value.

My two-sector model can be reinterpreted as a one-sector model with stochastic adjustment costs. In a one-sector model without adjustment costs the relative price of the investment good is always one, because investment can be transformed into consumption in a linear fashion. The introduction of convex adjustment costs imposes some curvature on the investment-consumption possibility frontier, since the marginal cost of investment rises with the rate of investment. In contrast, my model features a stochastic investment-consumption possibility frontier. Specifically, allocating more labor to the investment sector increases investment at the cost of consumption, and the tradeoff depends on the state of technology in the investment sector, $Y$.

**1.3.3 Households**

There exists a continuum of identical households that maximize expected utility of consumption,

$$J = \mathbb{E}_0 \int_0^\infty e^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma} dt.$$

(1.8)
Households supply a fixed amount of labor\(^1\) that can be freely allocated between the two sectors,

\[ L_{C,t} + L_{I,t} = 1. \]  
\[ (1.9) \]

Shifts in the allocation of labor between the two sectors allow the economy to intratemporally trade off consumption versus investment. An alternative interpretation of \( L \) is as a perishable good which can be used as input in either of the two sectors and which gets produced at a fixed flow, for example oil. A third interpretation is as a reduced form for government subsidies or tax exemptions that affect the productivity of each sector. The only requirement is that the factor be in short-run fixed supply and flexible between sectors. The model nests the case where there is no fixed factor when \( \beta_C = \beta_I = 1 \).

Households trade a complete set of state contingent securities in the financial markets. I focus on the empirically relevant case of \( \gamma > 1 \). Finally, the parameters in the model are assumed to satisfy

\[ u \equiv \rho - (1 - \gamma)(\mu_X - \delta_C) + \frac{1}{2} \gamma(1 - \gamma)\sigma_X^2 > 0 \]  
\[ (1.10) \]

The first restriction ensures that the value function for the social planner is bounded from below, whereas the second ensures that the state variables of the economy have a stationary distribution.

### 1.3.4 Competitive equilibrium

**Definition 1.** A competitive equilibrium is defined as a collection of stochastic processes \( C^*, K_C, I_C, L_C, \pi^*, \lambda^*, \omega^* \) such that (i) households chose \( C^* \) to maximize (1.8) given \( \omega^* \) and \( \pi^* \) (ii) firms choose \( I_C, L_C, \) and \( \lambda^* \), given \( \pi^*, \omega^* \) and \( \lambda^* \), to maximize (1.4) and (1.7) (iii) \( K_C^* \) satisfies equation (1.3) given \( I_C^* \) (iv) all markets clear according to (1.1), (1.5) and (1.9)

In this section I focus on the social planners problem, that is the problem of optimal allocation of labor. I demonstrate in the Appendix that, as in other models with dynamically complete financial markets and no externalities, a competitive equilibrium can be constructed from the solution to the social planner’s problem.

**Proposition 1.** The social planner’s value function is

\[ J(X, Y, K_C) = \frac{(X_t K_C^2)^{1-\gamma}}{1 - \gamma} f(\omega_t), \]

\(^1\)Endogenizing the supply of the fixed factor would introduce additional wealth and substitution effects in the model. If one keeps the labor interpretation, this would involve introducing labor in the utility function. The net effect on the marginal utility of consumption would depend on the form chosen and on the elasticity of substitution between consumption and leisure. Nevertheless, empirical evidence suggests that labor supply is quite inelastic, see for example Kimball and Shapiro (2003).
where \( f(\omega) \) satisfies the ODE

\[
0 = \left\{ L^*_{C,t} (1-\beta_C)(1-\gamma) + c^{-1} \left( \alpha e^\omega L^*_{I,t} (1-\beta_I) \right) (\beta_C (1-\gamma) f - f') - u f + (\mu_Y + \delta_C) f' \\
+ \frac{1}{2} \sigma^2_Y (f'' - f') \right\},
\]

and

\[
\omega_t \equiv \ln \left( \frac{Y_t}{K_{C,t}} \right).
\]

The allocation of labor between the two sectors is given by \( L^*_{C,t} = 1 - l(\omega_t) \) and \( L^*_{I,t} = l(\omega_t) \) where \( l(\omega) \) is defined by

\[
(1 - l(\omega))^{(1-\beta_C)(1-\gamma)-1} = l(\omega)^{\beta_I} \alpha \frac{1 - \beta_I}{1 - \beta_C} e^\omega \left( \beta_C f(\omega) - \frac{f'(\omega)}{1-\gamma} \right) c^{-1} \left( \alpha e^\omega l(\omega)^{1-\beta_I} \right).
\]

The state variable, \( \omega \), has dynamics

\[
d\omega_t = \left( \mu_Y + \delta - \frac{1}{2} \sigma^2_Y - \alpha e^\omega l(\omega_t)^{1-\beta_I} \right) dt + \sigma_Y dZ^Y_t.
\]

Proof See Appendix.

I solve for equilibrium policies numerically, and the details of the solution are shown in the Appendix.

In equilibrium, there is only one state variable that determines optimal policy, \( \omega \), and it can be interpreted as the ratio of "effective" capital stocks in the two sectors. Most importantly, innovations to \( \omega \) come only from the I-shock (Y). Finally, as long as \( \mu_Y + \delta - \frac{1}{2} \sigma^2_Y > 0 \), the state variable has a unique stationary distribution. This guarantees that in equilibrium one sector does not dominate the economy.

The mechanism that determines the price of risk for the I-shock is the allocation of labor between the sectors, \( l(\omega) \). If investors have time-separable preferences, only shocks that affect consumption contemporaneously are priced. Capital in the C-sector cannot instantaneously adjust in response to an I-shock. Therefore, the only mechanism through which consumption can instantaneously respond to the I-shock is through shifts in the allocation of labor between the sectors.

I find that \( l(\omega) \) is an increasing function of \( \omega \). This is important because it means that consumption temporarily falls after a positive I-shock, and that the I-shock has a negative premium. The behavior of \( l(\omega) \) can be understood from the first order conditions of the planner's problem evaluated at the equilibrium allocation. The first order conditions for
optimality are

\[ \pi_s = e^{-\rho_s} U_C, \]  
\[ \lambda_s = \frac{J_K}{U_C} c(i^*_C) = X K^{\beta_C-1} \frac{\beta_C(1 - \gamma)f - f'}{(1 - \gamma)(1 - l(\omega))\gamma(\beta_C-1)} \frac{1}{c'(ae^U(l(\omega)^{1-\beta_I}))}, \]  
\[ \lambda_Y \frac{\alpha K_I^{\beta_I}}{X K_C^{\beta_C}} = \frac{l(\omega)^{1-\beta_I}}{(1 - l(\omega))^{1-\beta_C}} \frac{1 - \beta_C}{1 - \beta_I}. \]

Here, \( \pi \) is the shadow cost of the resource constraint in the C-sector, (1.1), and \( \lambda \pi \) is the shadow cost of the resource constraint in the I-sector, (1.5). This implies that \( \pi \) is the state price density and that \( \lambda \) is the relative price of the investment good in terms of the consumption good.

Equation (1.11) is standard and states that, in equilibrium, the marginal utility of consumption in each state equals the shadow cost of the resource constraint in the C-sector, which is the state price density.

Equation (1.12) states that the relative price of output in the I-sector, \( \lambda \), equals the marginal value of capital in the C-sector divided by the marginal installation cost and by marginal utility. In the one sector model without adjustment costs, the relative price of the investment good is always one and marginal utility equals the marginal value of capital. In my model, \( \lambda \) is a function of the I-shock, because a positive investment shock increases the supply of the investment good and therefore lowers its relative price. In addition, \( \lambda \) depends on the N-shock (X), because a positive shock to productivity in the consumption sector increases the demand for the investment good and therefore its relative price. This is the reason why, in equilibrium, the N-shock affects both sectors symmetrically.

Equation (1.13) states that in equilibrium, the marginal product of labor in both sectors must be equal. This condition determines \( l(\omega) \).\(^2\) The effect of the investment-specific shock on the allocation of labor depends on how \( Y \) affects \( \lambda(Y)Y \). As long as \( \lambda(Y)Y \) is increasing in \( Y \), a positive shock to \( Y \) increases the profits of firms in the investment sector as well as the marginal product of labor in the I-sector. Therefore, the allocation of labor to the I-sector, \( l(\omega) \), must temporarily increase, inducing a fall in consumption. In the future, the capital stock in the consumption sector increases, reversing the short-run fall in consumption. The end result is that consumption displays a U-shaped response to a positive investment shock. Consumption initially falls, as more resources are allocated to the I-sector in order to take advantage of the improvement in technology. Eventually, the new technology starts bearing fruit and the growth rate of consumption increases.

Finally, the pricing kernel is

\[ \frac{d\pi_t}{\pi_t} = -r_t dt - \gamma s_X dZ_t^X - \gamma s(\omega) dZ_t^Y, \]

\(^2\)This also illustrates one of the reasons why the model does not admit an analytic solution, since \( l(\omega) \) cannot be determined in closed form.
where \( s(\omega) \equiv -(1 - \beta C) \frac{f'(\omega)}{1 - f(\omega)} \).

### 1.3.5 Asset prices

#### Investment and consumption firms

In the model there are two representative firms, one producing the consumption good and one producing the investment good. The market value of each firm is

\[
S_C^t = E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( X_s K^\beta_C (L^*_C, s)^{1-\beta C} - w_s L^*_C, s - \lambda_s C \left( \frac{I^*_C, s}{K^*_C, s} \right) K^*_C, s \right) ds, \quad (1.15)
\]

\[
S_I^t = E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( \lambda_s \alpha Y_s K^\beta_I (L^*_I, s)^{1-\beta I} \right) ds. \quad (1.16)
\]

The value of each sector includes all cashflows accruing to the owners of the capital stock.

**Proposition 2.** The ratio of the value of the investment goods sector, \( S_I^t \), over the consumption goods sector, \( S_C^t \), equals

\[
\frac{S_I^t}{S_C^t} = \frac{\beta I f'(\omega)}{\beta C (1 - \gamma) f(\omega) - f'(\omega)},
\]

and is an increasing function of \( \omega \).

**Proof** See Appendix.

A positive I-shock increases the productivity of the investment sector and increases the value of the investment sector relative to the consumption sector. This has two important implications.

The first is that the relative value of the two sectors is a state variable in the economy, since it is a monotone transformation of \( \omega \). The same is not true for the relative price of the investment good, \( \lambda \), which also depends on the N-shock, as shown in equation (1.12). Therefore, the cross-section of stock prices may contain additional information about real investment opportunities relative to the prices of investment goods or Tobin’s Q. This information can be used to identify investment-specific technological change in the data.

The second implication is that a portfolio of investment minus consumption stocks (IMC) is positively correlated with the I-shock. Hence, I can use returns on the IMC portfolio to test the asset pricing implications of the model, namely that the pricing kernel loads on the I-shock with a negative premium. If the model is true, then the IMC portfolio should have negative expected returns after adjusting for market risk. Additionally, stocks that load positively on the IMC portfolio should have lower expected returns than stocks with low loadings on IMC. Furthermore, if the IMC portfolio spans a systematic source of risk, it should be able to explain the variation of realized returns.
Value versus growth

Because markets are complete, claims to cashflows can be decomposed in such a way as to create the analog of value and growth firms. The value of a firm can be separated into the value of assets in place and the value of future growth opportunities.\(^3\) When investment is frictionless, existing assets represent the entire value of the firm, since a perfectly elastic supply of capital means that all future projects have zero NPV. In contrast, frictions in investment prevent the supply of capital from being perfectly elastic, hence future growth opportunities have additional value.

I focus on the value of assets in place and growth opportunities in the C-sector, since in the baseline model the capital stock in the I-sector is fixed.\(^4\) The value of assets in place in the consumption sector equals the value of all dividends accruing from existing assets

\[
S_t^V = \max_{L_{C,t}} E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( X_s (K_{C,t} e^{-\delta(s-t)})^{\beta_C} \hat{L}_{C,s}^{1-\beta_C} - w_s \hat{L}_{C,s} \right) ds.
\]

(1.17)

In the absence of arbitrage, the value of growth opportunities must equal the residual value

\[
S_t^G = S_t^C - S_t^V.
\]

(1.18)

This decomposition creates fictitious value and growth firms in the economy without the cost of modeling individual firms explicitly.

**Lemma 1.** The relative value of assets in place over growth opportunities in the consumption sector equals

\[
\frac{S_t^V}{S_t^G} = \frac{g(\omega)}{\beta_C (1-\gamma) f(\omega) - f'(\omega) - g(\omega)},
\]

where the function \(g(\omega)\) satisfies the ODE

\[
0 = \beta_C (1 - \gamma) L_{C,s}^{1-\beta_C (1-\gamma)} - u g(\omega) + i^*_C (\beta_C (1 - \gamma) - 1) g(\omega) + D_\omega g(\omega).
\]

**Proof** See Appendix.

Lemma 1 implies that innovations to \(S_t^V / S_t^G\) are independent of the neutral productivity shock, \(dZ^X\), and are spanned by the investment-specific shock, \(dZ^Y\).

I find that the relative value of assets in place minus the value of growth opportunities in the consumption sector is decreasing in \(\omega\), as shown in figure 5(i). This is important because it offers a novel explanation for the value effect. A positive I-shock lowers the cost of new capital in the C-sector, increasing the value of future growth opportunities relative to the value of installed assets. Therefore, since the I-shock carries a negative premium,

\(^3\)An early example of work that decomposes firm value into value of assets in place and future growth opportunities to explore the value puzzle is Berk, Green and Naik (1999).

\(^4\)In the Appendix, I solve the general model with two capital stocks and compute the value of assets in place and growth opportunities for the entire economy. The results are qualitatively and quantitatively comparable.
this implies that growth stocks have lower average returns than value stocks. In figure ??, I plot the risk premium on the portfolio of fictitious value minus growth firms as a function of different parameters. For all parameter values considered the premium is positive.

In addition to explaining the value premium, my model can explain the comovement of value and growth firms. One of the major contributions of Fama and French (1993) is to show that value and growth firms move together, or equivalently that the HML portfolio can explain the cross-section of realized as well as expected stock returns. Lemma 1 implies that a portfolio of value minus growth stocks also spans the investment-specific shock, creating comovement between value and growth firms. Conversely, models that explain the value premium with only one aggregate shock, cannot generate comovement in value and growth firms independent of the market portfolio. The presence of a second aggregate shock highlights one crucial difference between this chapter and Gomes, Kogan and Zhang (2003) and Gala (2006) who argue that value firms are riskier than growth firms in bad times, and hence a conditional CAPM should price the value spread. In this chapter, the value premium arises due to exposure to a second aggregate shock, the I-shock, and therefore the conditional CAPM does not hold.\footnote{However, as in any model with time-separable preferences, the CCAPM does hold conditionally. Nevertheless, recent research has shown that the CCAPM is somewhat successful at explaining the book to market cross section (Lettau and Ludvigson (2004), Jagannathan and Wang (2005), Parker and Julliard (2005), Bansal, Dittmar and Lundblatt (2005), Santos and Veronesi (2006a), Malloy, Moskowitz and Vissing-Jorgensen (2006) and Yogo (2006) among others).}

The decomposition of aggregate value into value of assets in place and future growth opportunities though stylized, helps focus on the direct effects of investment shocks on aggregate dynamics, instead of on indirect effects through the aggregation heterogenous firms. A potentially interesting extension would explore firm-level investment decisions in more detail by incorporating firm level productivity shocks and possibly financial frictions.

**Market portfolio and the riskless asset**

The sum of the market values of the two sectors equals the market portfolio or the value of the entire dividend stream

\[
S_t^M = S_t^C + S_t^I = E_t \int_t^\infty \frac{\pi_s}{\pi_t} (C_s - w_s) \, ds
\]

\[
= \frac{(X_t K_{C,t})^{1-\gamma}}{1-\gamma} (\beta_C (1-\gamma) f(\omega) + (\beta_I - 1) f'(\omega)).
\]  

(1.19)

Returns on the market portfolio are driven by innovations in the neutral shock, \(dZ^X\), and the investment shock, \(dZ^Y\). When solving the model, I find that the value of the market portfolio is positively correlated with the neutral shock and negatively correlated with the investment shock. A positive I-shock increases future dividends and decreases future state prices. If agents have CRRA preferences with \(\gamma > 1\), the effect on state prices dominates and leads to a fall in the price of the dividend stream. Given that the I-shock has a negative
price of risk, this increases the equity premium.

As in any equilibrium model with time-separable preferences, the rate of return on the safe asset is a function of the first two moments of consumption growth. The riskless rate equals

\[ r_t = \rho + \gamma m(\omega) - \frac{1}{2} \gamma (\gamma + 1) (s(\omega)^2 + \sigma^2), \]

where \( s(\omega) = -(1 - \beta C) \frac{y'(\omega)}{1-\ell(\omega)} \), and \( m(\omega) \) is the drift of the optimal consumption process defined in the Appendix.

Households want to smooth consumption across time, so when consumption growth is high they borrow against the future, leading to higher interest rates. Households also have a precautionary savings motive, which induces them to save more when their consumption stream is more volatile. Investment in the consumption sector is increasing in \( \omega \), hence the growth rate of consumption is also increasing in \( \omega \). Given that the precautionary savings effect is quantitatively small, the riskless rate is an increasing function of \( \omega \) and inherits its mean reverting properties.

### 1.4 Empirical evidence

The model in the previous section has a number of empirical implications about the cross-section of stock returns. I provide evidence that these implications are supported by the data.

The first implication is that investment-specific technological change earns a negative risk premium. To see this, note that the pricing kernel implied by the model can be linearized as

\[ \pi = a - b_X dZ_X - b_Y dZ_Y. \]

This is the empirical equivalent of equation (1.14) linearized around \( E(\omega) \). The model implies that shocks to investment technology, \( dZ_Y \), have a negative price of risk, or equivalently that \( b_Y < 0 \). A positive investment-specific shock lowers the cost of new capital and thus acts as a shock to the real investment opportunities in the economy, increasing the marginal value of wealth and therefore state prices. As a result, firms that covary positively with the investment-specific shock should have, ceteris paribus, lower average returns.

The second implication is that a portfolio that is long the value of assets in place and short the value of growth opportunities is negatively correlated with the I-shock, and should therefore earn a positive premium. To the extent that value firms have more assets in place and fewer growth opportunities than growth firms, sorting stocks on book to market should produce a positive spread in returns that is explained by their covariance with a proxy for the I-shock. Sorting stocks on book to market is well known to produce portfolios that are mispriced by the CAPM. This is the well documented value puzzle.

However, in order to test the empirical implications of the model, it is necessary to identify investment-specific technological change in the data. One of the advantages of the
model is that it provides restrictions that identify investment-specific shocks. As shown in Propositions 1 and 2, random fluctuations in the ratio of the value of the investment goods sector, \( S^I_t \), to that of the consumption goods sector, \( S^C_t \), are driven only by the investment-technology shock, \( dZ_t^X \), and are unrelated to the neutral productivity shock, \( dZ_t^X \). Therefore, the random component of \( dY_t/Y_t \) is spanned by a zero-investment portfolio that is long the investment goods sector and short the consumption goods sector. Thus, the cross-sectional implications of risk exposure to the investment shock can be investigated by including returns to this portfolio in standard factor pricing models. This enables me to identify innovations in investment-specific shocks by the return on an investment minus consumption portfolio.

1.4.1 Investment minus consumption portfolio

Ideally, the distinction between firms producing investment and consumption goods would be clear and the new factor portfolio would be straightforward to obtain. However, many companies produce both types of goods. In order to overcome this difficulty, I construct two proxies by classifying industries as consumption or investment good producers\(^6\). I use information from the US Department of Commerce’s NIPA tables with the procedure described in the Appendix. The first approach classifies industries based on the sector they contribute the most value and obtains the first proxy, IMC. The second proxy, denoted by IMCX, is constructed by only including industries that contribute to one sector and not the other. Both proxies have relative advantages and disadvantages. While the IMC portfolio maximizes the fraction of market value that is classified, it can potentially misclassify industries because it relies on measures of value. On the other hand, the IMCX portfolio minimizes this misclassification but results in a lower fraction of market value covered.

The composition details are displayed in table 1. With either definitions, the sector producing consumption goods is much larger than the sector producing investment goods, both in number of firms and in term of market capitalization. Further, with either classification schemes, the consumption and investment portfolio have fairly similar ratios of book to market equity. As a robustness check, I construct an investment minus consumption portfolio using the data provided by Gomes, Kogan and Yogo (2006). I create this portfolio, labeled \( IMC_{GKY} \), by subtracting from the investment portfolio a market-capitalization weighted average of the services and the non-durables portfolios. The correlation of the IMC(X) portfolios with \( IMC_{GKY} \) and the Fama-French factors is displayed in table 2. All three investment minus consumption portfolios are highly correlated, especially IMC and \( IMC_{GKY} \). In addition, the IMC portfolios have negative correlation with HML and small but positive correlation with the market. The negative correlation with HML is consistent with the model, since the value of assets in place minus growth opportunities is negatively correlated with the I-shock. Finally, the IMC(X) portfolio has a positive correlation with the SMB factor, which stems from the fact that investment firms tend to have smaller market capitalization.

\(^6\) A similar procedure is followed by Chari, Kehoe, and McGrattan (1996), Castro, Clementi and MacDonald (2006) and Gomes, Kogan and Yogo (2006)
Table 3 shows average returns on the three IMC portfolios along with their CAPM alpha and their correlation with HML, broken down by decade. The IMC portfolios have negative average returns, and a negative CAPM alpha. This is consistent with the results of Gomes, Kogan and Yogo (2006) and Makarov and Papanikolaou (2006). More importantly, the IMC portfolios have higher correlation with HML in the later half of the sample, during which average returns and CAPM alphas on IMC and HML are higher in magnitude. This evidence is consistent with the dramatic surge in innovations in investment goods over the last few decades, such as personal computers and the internet. An increase in the volatility of the investment shock would translate into a higher premium for HML and IMC and would increase the correlation between the two.\footnote{In the model, IMC and HML are perfectly correlated, because they are both spanned by the investment-specific shock. In the presence of additional aggregate shocks, this need no longer be the case. Nevertheless, an increase in the volatility of the investment shock would increase the correlation between IMC and HML, which is consistent with the evidence in the data.}

The model implies that investment-specific technological change is a source of systematic risk. If IMC proxies for investment-specific shocks, then it must also explain variation in realized stock returns. Fama and French (1993) show that HML and SMB are responsible for a significant amount of variation in the realized returns of book to market and size sorted portfolios. This is important because it provides evidence that HML and SMB are risk factors rather than simple manifestations of mispricing. If value and growth stocks were simply mispriced by investors, it does not immediately follow that they would move together.

I compute the total variation of realized stock returns that can be attributed to IMC, and contrast with how much variation is explained by the market portfolio and the Fama-French book-to-market factor, HML. For every common stock in the CRSP database that is traded in NYSE, AMEX or Nasdaq, I estimate the following regression

\begin{equation}
\begin{align*}
    r_{it} &= a + \beta_t f_t + u_t,
\end{align*}
\end{equation}

where $r_{it}$ are log returns of the individual stock and $f_t$ are log returns on factors (log excess returns on the market portfolio, HML or IMC). I compute the adjusted $R^2$ for each regression. I report the average adjusted $R^2$ across stocks, both equally-weighted, and weighted by the variance of each individual stock.\footnote{Since the $R^2$ is the ratio of explained over total variance, the first measure is simply the sum of ratios across stocks, whereas the second measure is the ratio of the sums.}

Table 4 reports the results. The market portfolio explains 12.0-13.7\% of the total variation in individual stock returns. When IMC or HML is added to the regression, the average $R^2$ increases by 1.6-2.0\%. Interestingly, IMC and HML increase the total explained variation by a comparable amount. However, if both portfolios are added in the specification, the average $R^2$ increases by an additional 0.7-1.0\%, suggesting that each portfolio may contain some information not captured by the other.
1.4.2 Cross-sectional tests

If I-shocks are spanned by the investment minus consumption portfolio and earn a negative premium, then sorting stocks on their covariances with IMC into portfolios should produce a negative spread in expected returns that is not explained by the market portfolio. Moreover, the IMC portfolio should be able to price this spread. Table 5 presents summary statistics on 10 portfolios of stocks constructed by sorting on covariances with IMC. This sort produces an almost monotone decline in average returns and a spread of roughly 2.5% annually. However, it is more informative to consider CAPM alphas because the portfolios have different risk profiles, as evidenced by the increasing pattern of both the standard deviation and the market beta of each portfolio. The pattern displayed by the pricing errors is more striking. The pricing errors of the portfolios decline almost monotonically from 2.6% to -2.4% annually. The pattern is dampened but not eliminated if one looks at alphas from the Fama-French three factor model. Finally, including the IMC portfolio in the CAPM significantly reduces the pricing errors. The GRS F-test rejects both the CAPM and the three factor model at the 1% level, whereas it fails to reject the model with the market and IMC portfolios at the 10% level.

The pricing kernel in equation (1.20) summarizes all the cross-sectional asset pricing implications of the model. The restrictions on the rate of return of all traded assets that is imposed by no arbitrage,

\[ E[\pi R] = 1, \quad (1.22) \]

can be used to estimate (1.20) by the generalized method of moments. Accordingly, I estimate the model using two-stage GMM with the details described in the Appendix. I report the mean absolute pricing error (MAPE), the sum of squared pricing errors (SSQE) and the J-test of the over-identifying restrictions of the model, namely that all the pricing errors are zero. I use returns on the CRSP value-weighted portfolio and monthly non-durable consumption growth from NIPA as empirical proxies for the N-shock and focus on the period 1959-2005. I use the return on the IMC portfolio as proxy for the I-shock. I compare the performance of the two models incorporating IMC with the CAPM, the CCAPM and the Fama-French three factor model. Finally, the model implies that HML derives its pricing ability through its exposure to the investment shock. I therefore include both HML and IMC in the same specification in order to see if each factor has additional pricing ability in the presence of the other. Because estimating (1.20) using the entire cross-section of stock returns can be problematic, the literature focuses on a particular subset of assets, which are portfolios of stocks sorted on economically meaningful characteristics.

This chapter examines on whether investment-specific shocks are an important component of the pricing kernel. I therefore focus on the estimate of \( b_Y \), rather than on the overall

---

9The estimated parameters from the first and second stage are qualitatively and quantitatively similar.

10I choose to report the sum of squared errors rather than the normalized sum of squared errors (\( R^2 \)) because in the absence of a constant term it is not clear what the benchmark (i.e. the normalization term) is.

11Results using the IMCX or IMC_GKY portfolios are qualitatively and quantitatively very similar and are available upon request.
ability of the model to price each cross-section, which might depend on the particular choice of proxy for the neutral productivity shock. Most importantly, because (1.20) must hold for all traded assets in the economy, estimates of \( b_\gamma \) should be robust to using different test assets. To this end, I estimate (1.20) using a number of different cross-sections.

**Risk-sorted portfolios**

I consider as test assets portfolios of stocks sorted first by industry, market beta or market capitalization, and then covariance with the IMC portfolio. The construction of these portfolios is standard and is described in the Appendix. I estimate (1.20) using the above set of test assets.

Table 6 reports estimates using 24 portfolios sorted first on industry and then on IMC covariance. The CAPM and (C)CAPM perform rather poorly in this cross-section, with risk prices on the market portfolio and consumption growth not statistically different from zero. Both models generate large absolute pricing errors, though the J-tests fail to reject each model due to the large standard error of the pricing errors. The Fama-French three factor model performs significantly better, with smaller absolute pricing errors and with half the variance of pricing errors. However, this improvement is mostly driven by the negative and statistically significant premium on SMB, whereas the premium on HML is positive but not statistically significant. The two models that include the IMC portfolio are substantially more successful than the (C)CAPM. The estimated risk prices on IMC are negative and statistically different from zero. The estimated risk price on the market is also positive and significant, but the same is not true for the price of consumption growth. Both models produce lower absolute pricing errors and smaller variation of pricing errors than the Fama-French three factor model, even though they have fewer number of factors. The last column includes both IMC and HML along with the market portfolio. The estimated premium on IMC is still negative at -3.71 and -4.39 respectively and statistically significant from zero, whereas the premium on HML is negative and not statistically significant.

Tables 7 and 8 repeat the analysis using 25 portfolios sorted first on market betas and market capitalization, and then on covariance with IMC. The results are similar in all three cross-sections. The CAPM and CCAPM perform rather poorly, producing large pricing errors both in terms of absolute magnitude and in terms of variation across assets. Including the IMC portfolio in the (C)CAPM significantly improves the performance of the model, making it comparable to the Fama-French three factor model in terms of pricing errors. The estimated premium on IMC is always negative, statistically significant and in the -2.6 to -4.3 range. The same is not true however for the estimated premia on SMB and HML, which in most cases are not statistically significant from zero. Finally, including HML and IMC in the same specifications results in an estimated price of risk for both factors that is not

\[^{12}\text{Nagel, Lewellen and Shanken (2006) caution that if the cross-section of test assets has a strong factor structure, then some factors may appear to price the test assets arbitrarily well if they are correlated with the common factors. This is an argument against evaluating an asset pricing model only on a set of portfolios with a strong factor structure (i.e. the 25 BM/ME) or portfolios that are only sorted on covariances with the proposed factor.}\]
The main results of this section can also be graphically displayed. Figures 2(b), 2(c) and 2(d) plot the CAPM pricing errors for the above sets of portfolios against their covariance with the IMC portfolio. One can see that all sorts produce a significant spread in pricing errors of around 6%-8% annually, and that there is a strong negative relationship between pricing errors and covariances with the IMC portfolio. This suggests that the CAPM omits a significant source of systematic risk, specifically risk associated with the IMC portfolio. Most importantly, the relationship is negative, which is consistent with a negative premium on I-shock.

Book to market sorted portfolios

Table 9 presents estimation results for the 25 Fama-French portfolios sorted on book to market and market capitalization. As is well known, the CAPM performs poorly in pricing this cross-section, generating large pricing errors, both in terms of absolute magnitude and variation across assets. Even though the estimated price of risk on the market portfolio is statistically significant, the mean absolute pricing error is 2.8%. The CCAPM is performing somewhat better with a mean absolute pricing error of 2.2%. The estimated price of risk on consumption growth is positive, statistically significant and equal to 74.7, with the magnitude consistent with the literature on the equity premium puzzle. The Fama-French three-factor model performs significantly better, with the sum of squared errors equal to 0.65 vs 3.01 for the CAPM and 1.89 for the CCAPM.

Including the IMC portfolio in the (C)CAPM dramatically improves the performance of both models. More importantly, the estimated premium on the IMC portfolio is negative at -4.04 and -5.35 respectively, and statistically different from zero. The overall performance of the model however depends on the choice of proxy for the N-shock. The model that includes the market portfolio does substantially worse, with the sum of squared errors equal to 2.22. This is partially due to the fact that IMC cannot explain the size premium. If SMB is added to the specification, then the model performs as well as the Fama-French three factor model with a sum of squared errors of 0.65, but with a substantially higher estimated premium on IMC at -10.8. The model with only consumption growth and IMC performs slightly better than the Fama-French model, generating a sum of squared errors of 0.56 with only two factors. Finally, including the IMC portfolio in the Fama-French model does not improve the performance of the model, with the estimated price of risk on IMC still negative but not statistically significant. The results of this section suggest that the IMC portfolio embodies similar pricing information as HML. This is consistent with the model, because HML can also be used as a proxy for investment-specific shocks.

Table 10 reports estimation results using the 25 BM/IMC covariance sorted portfolios. The CAPM and CCAPM perform rather poorly, generating mean absolute pricing errors of 2.3%-2.4% and sum of squared errors of 2.22 and 1.87 respectively. The Fama-French model performs significantly better, generating a mean absolute pricing error of 1.2% and a sum of squared errors of 0.57. As before, including the IMC portfolio in the (C)CAPM dramatically improves the pricing performance of the model, reducing the variation in pricing errors by
half to 1.08 and 0.87 respectively. Again, the estimated premium on IMC is negative at -3.45 and -3.95 and statistically significant.

The main results of this section can also be graphically displayed. Figures 2(e) and 2(f) plot the CAPM pricing errors for the above sets of portfolios against their covariance with the IMC portfolio. As before, there is a negative relationship between the pricing errors and covariances with IMC, suggesting that the CAPM omits a systematic source of risk that may be proxied by the IMC portfolio.

Individual stocks

In this section I repeat the cross-sectional tests using the entire cross-section of returns. The problem when using individual stocks is that covariances are measured with error, which biases the estimated risk premium towards zero. To alleviate measurement error I follow the procedure of Fama and French (1992) of first sorting stocks into portfolios and then assigning the portfolio covariance with IMC to the individual stocks. The full procedure is described in the Appendix.

The estimation results are shown on table 11. The first panel presents results for the entire sample. The estimated premium on IMC is -3.69 and statistically significant, even when the log book to market ratio and the log market capitalization is included in the specification. Most importantly, the estimate of $b_y$ is very close to the estimates obtained using different sets of test assets.

In the second and third panels of table 11, I present estimates using the first and second half of the sample. I find that the estimated premium on IMC is -5.45 in the second half of the sample and -1.91 in the first half, with t-statistics of -2.59 and -1.26 respectively. On the one hand, this is consistent with the negative premium on IMC being significantly stronger in the second half of the sample, as suggested by table 3, and with the fact that the rate of technological innovation in investment goods has been significantly higher in the last 25 years. On the other hand, care must be taken when interpreting the sub-sample results. The reason is that during the first sub-sample the number of stocks in each portfolio is much smaller, so each portfolio is less diversified. It is very likely that measurement error is higher during the first sub-sample, which would artificially make the premium smaller.

1.4.3 Does IMC proxy for investment-specific technological change?

Proposition 2 identifies the IMC portfolio as a proxy for the investment-specific shock. However, it is possible that this portfolio captures other sources of systematic risk or perhaps proxies for the systematic mispricing of some stocks. If the IMC portfolio is a valid proxy for investment-specific shocks, then it must satisfy the following.

First, IMC must be a factor that drives the time series of realized returns. If stocks A and B have high average returns because they are exposed to similar risk, then they must also move together. Makarov and Papanikolaou (2006) show that one of the factors driving the cross-section of industry portfolios is highly correlated with a portfolio of investment minus consumption goods industries. In addition, the results from table 4 suggest that the IMC
portfolio captures a significant amount of variation in realized returns that is quantitatively comparable with HML.

Second, there must be a link between IMC and aggregate price and quantity of investment. Specifically, if this shock captures investment-specific productivity shocks, then it must predict an increase in investment and a fall in the quality-adjusted price of the investment good.

Third, if IMC is part of the pricing kernel, there must be some a link, at some horizon, between cashflows or returns on this portfolio and aggregate consumption. Even though in the model consumption can adjust instantaneously via the allocation of labor across sectors, this is a simplification. In reality, it might take several quarters for consumption to adjust. If the IMC portfolio captures an investment-specific shock, it should predict a short run fall and a long-run increase in the growth rate of consumption.

**Response of investment**

I explore how investment and the price of new investment goods respond to returns on the IMC portfolio. In particular, I look at investment in non-residential structures and equipment and software, and decompose the latter into investment in information equipment and investment in industrial equipment. I use the investment quantity indices from BEA. Regarding the price of new equipment, quality-adjustment is an issue. The reason is that if investment-specific shocks also represent an increase in the quality of the investment good, then its price might increase. This makes it difficult to disentangle quality-improving productivity shocks from demand-side shocks, because both could predict an increase in prices and quantities. Unfortunately, obtaining quality adjusted-prices of investment goods is somewhat problematic. According to Moulton (2001), NIPA currently incorporates hedonic methods to quality-adjust computers, semiconductors and digital telephone switching equipment, but other types of equipment deflators are not adjusted. Consequently, I use the price index for computers and peripheral equipment relative to the GDP deflator as a proxy for the quality-adjusted price for new equipment.

I estimate

\[ x_{t+k} - x_t = \alpha_0 + \beta_k R_{IMC,t} + \gamma_k R_{MKT,t} + \epsilon_t, \]

where \( x_t \) denotes log investment or the log relative price deflator at time \( t \), and \( R_{IMC,t} \) and \( R_{MKT,t} \) denote returns on IMC and the market portfolio respectively. The IMC portfolio is correlated with the market portfolio, which is known to predict investment and consumption. Therefore I control for returns to the market portfolio.

Figure 1-3 plots the \( \beta_k \) coefficients along with 10% confidence intervals using HAC standard errors adjusted by Newey-West. I find that investment in equipment and non-residential structures sharply increases following positive returns on the IMC portfolio. The increase in investment following positive returns on IMC is statistically significant for up to 5-6 quarters forward. At the same time, the quality adjusted price of equipment falls following positive returns on IMC, but the response is statistically significant for only one quarter ahead. Investment in non-residential structures also displays a strong response, which is consistent
with the findings of Gort, Greenwood and Rupert (1999) who document significant technological advances in structures.

**Response of consumption**

In any equilibrium model where agents optimize their consumption, factor risk premia must be ultimately linked to how consumption covaries with the factor. I look at the cumulative response of consumption on returns to the IMC portfolio, controlling for the market portfolio

\[ x_{t+k} - x_t = \alpha_0 + \beta_k R_{IMC,t} + \gamma_k R_{MKT,t} + \epsilon_t, \]

where here \( x_t \) denotes log consumption. I focus on the discretionary component of consumption. I employ the quantity indices from BEA on consumption expenditures excluding food and energy, consumption of durable goods and residential investment. Even though the latter is formally part of the investment sector, I consider residential investment as a durable good that provides a constant flow of consumption services. Finally, I also consider the total output of the economy, as measured by GDP minus government expenditures.

Figure 1-4 presents the estimated \( \beta_k \) coefficients along with 10% confidence intervals using HAC errors adjusted by Newey-West. For all three measures of discretionary consumption there is a similar U-shaped pattern. Consumption sharply falls following positive returns on the IMC portfolio. This fall is reversed after 10-14 quarters, and the reversal is stronger for consumption excluding food and energy. Output on the other hand does not display a statistically significant response. Even though the short-run fall in consumption is statistically significant, the long run increase is not, partially because the number of independent observations shrinks rapidly with the horizon.

The pattern is consistent with the spirit of the model. A positive shock to investment technology does not affect output immediately, but it leads to a gradual reallocation of resources from the sector producing consumption goods to the sector producing investment goods. The increase in investment will eventually be reflected in an increase in output and consumption growth.

In order to overcome the difficulties with identifying long-run responses of consumption, I use the information embedded in real interest rates. Most equilibrium models have the implication that the expected real rate of return on the safe asset is an increasing function of consumption growth. The same is true in my model, as shown in section 3.5.3. Unfortunately, the real rate of return on the riskless asset is unavailable, because inflation-indexed bonds started trading in the US in 1997, while data on inflation expectations are only available starting from the late 70’s. To overcome this problem, I subtract realized inflation, as measured by the CPI-U index, from the nominal interest rate on the US Treasury Bonds. As long as investors have rational expectations, this variable could be a valid proxy for real rates.\(^{13}\) Additionally, Proposition 2 shows that the ratio of market values of the two

\(^{13}\)To see this note that subtracting realized inflation from the left hand side results in

\[ r_{t,t+\tau}^n - E_t(\pi_{t,t+\tau}) - (\pi_{t,t+\tau} - E_t(\pi_{t,t+\tau})) = a_0 + a_r \hat{w}_t + \epsilon_t. \]
industries is an increasing function of \( \omega \). This suggests employing the ratio of stock market capitalizations of the investment versus the consumption industry as a proxy for \( \omega \).\(^{14}\)

Since all variables are close to non-stationary, I estimate the following equation both in levels and in first differences

\[
r_{t,t+\tau}^n - \pi_{t,t+\tau} = a_0 + a_\tau \hat{\omega}_t + \epsilon_t,
\]

where \( r_{t,t+\tau}^n \) is the nominal yield on a Treasury Bond with maturity \( \tau \), \( \pi_{t,t+\tau} \equiv \ln CPI_{t+\tau} - \ln CPI_t \) is realized inflation, and \( \hat{\omega}_t \) is the ratio of market capitalizations of the two sectors. In addition, I estimate the equation replacing \( \Delta \hat{\omega} \) with returns on the IMC portfolio.\(^ {15}\)

The model implies that movements in the real risk-free rate are entirely driven by \( \omega \). In practice, of course, factors such as monetary policy could also be affecting real rates, and might be correlated with the relative value of the two industries. In order to address this issue, I consider two maturities, the 1 year and the 10 year rate. An investment-specific shock will affect consumption growth in the long run and should affect the 10 year rate more strongly than the 1 year rate. On the other hand, there is no reason for monetary policy to affect the long end more than the short end of the curve.\(^ {16}\)

Table 12 reports the estimates \( \hat{a}_1 \) along with t-statistics based on HAC standard errors. I present results using both classification schemes, estimated both in levels, in first differences and also employing returns on the IMC portfolio. The first panel shows the estimates for the 10-year rate. The estimated coefficient \( \hat{a}_{10} \) is positive, as the model predicts, and statistically significant regardless of the classification scheme and of whether the regression is estimated in levels, first differences or using returns on IMC. The bottom panel reports results for the 1-year rate. The estimated coefficient \( \hat{a}_1 \), even though positive, is in most cases not statistically different from zero.

The results presented in table 12 show that innovations in \( \hat{\omega} \), or equivalently returns on the IMC portfolio, are correlated with changes in the long-run real rates but not with changes in the short-run rates. This suggests that the IMC portfolio is more likely to capture changes in the long-run growth rate of consumption rather than changes in monetary policy. This, taken together with the evidence that consumption falls in the short run following positive returns on the IMC portfolio, is supportive of the model. A positive investment-specific shock leads to a reallocation of resources from consumption to investment, causing a temporary fall in consumption but an increase in its long-run growth rate.

\(^{14}\)This ratio is an imperfect proxy, because it does not include the value of debt issued by firms in both sectors and the value of private firms. However, the effect of measurement error will be less severe when the above regression is estimated in differences rather than levels.

\(^{15}\)Changes in the ratio of log capitalization are essentially the capital gains portion of returns on IMC. Unsurprisingly, the two are highly correlated.

\(^{16}\)If anything, the opposite may be true. Piazzesi (2005) shows that the estimated response of yields to monetary policy shocks declines with the maturity of the yield.
IMC and firm-level investment decisions

If the IMC portfolio captures investment-specific technological innovations, then positive returns on the IMC portfolio should predict also firm-level investment decisions. However, it is not clear that an investment shock will benefit all firms in the cross-section equally. Firms might be in different lines of business, or firms might be financially constrained and thus unable to benefit from technological innovations. This heterogeneity however, should be reflected in a firm’s stock returns, so if a firm benefits from an investment shock it should covary positively with the IMC portfolio.

I estimate the following regression using OLS

\[ \frac{I_{i,t+1}}{K_{i,t}} = a_i + a_1 \beta_{i,t}^{IMC} + a_2 R_t^{IMC} + a_3 \beta_{i,t}^{IMC} \times R_t^{IMC} + a_4 q_{i,t} + a_5 q_{i,t} \times R_t^{IMC} + u_{i,t}, \] (1.26)

where \( \beta_{i,t}^{IMC} \) is firm i’s beta with respect to IMC estimated using weekly observations at time t, \( R_t^{IMC} \) is returns on the IMC portfolio at time t, \( q_{i,t} \) is the logarithm of firm i’s equity market-to-book ratio at the end of year t, \( I \) is Capital Expenditures, \( K \) is Property, Plant and Equipment or Total Assets, \( a_i \) is the firm fixed effect. I also estimate the above regression by replacing \( R_t^{IMC} \) with a time fixed effect. To capture the idea that firms do not all benefit equally from investment-technology shocks, I interact \( \beta_{i,t}^{IMC} \) and \( q_{i,t} \) with \( R_t^{IMC} \). Firms that covary more with this shock (i.e., have higher \( \beta_{i,t}^{IMC} \)) should invest more after a positive shock has occurred. Responses to this shock should also depend on firms’ growth opportunities, proxied by \( q_{i,t} \). Standard errors are clustered at the firm level.

Table 1.6 presents the results. The coefficient on \( \beta_{i,t}^{IMC} \) is positive and statistically significant which indicates that firms with a higher covariance with the IMC portfolio invest more on average. Furthermore, firms invest more following a positive return on the IMC portfolio, since \( a_2 \) is positive and statistically significant. Most importantly, this relationship is stronger for firms that covary more with the IMC portfolio since the coefficient on the interaction term, \( \beta_{i,t}^{IMC} \times R_t^{IMC} \), is positive and statistically significant. This pattern is robust to the normalization of capital expenditures by PPE or Total Assets, and the use of time fixed-effects in place of \( R_t^{IMC} \) in the non-interaction term. However, it is possible that this result is simply capturing the fact that firms in different industries have different betas with respect to IMC and thus invest differently, rather than capturing variation within industries. To address this, I estimate equation 1.26 by also including industry-year fixed effects. The coefficient \( a_3 \) remains statistically significant and also increases in magnitude. This implies that it also captures variation within industries.

The results from this section, combined with the results from 4.2, raise the question of what determines the covariance of a firm’s stock returns on the IMC portfolio. Given that there is substantial variation in firm betas with IMC within an industry, and this variation is also associated with different investment policies, it is possible that not all firms within the same industry are able to benefit from a positive investment shock. It is possible that financial market imperfections, such as financial constraints prevent some firms from fully benefiting from technological innovations. A more detailed exploration of the determinants of firms’ loadings on investment-specific shocks and the role of financial constraints is left
for future research.

1.5 Conclusion

I extend the standard general equilibrium models used in finance to incorporate investment-specific technological change. Investment-specific shocks act as a shock to the real investment opportunities in the economy. In equilibrium, an increase in productivity of the capital goods sector is followed by a reallocation of resources from consumption to investment. This results in a negative price of risk for investment-specific shocks. In addition, the investment-specific shock affects the relative cost of new capital as well as the relative profitability of firms. In particular, a positive investment-specific shock increases the value of investment good producers and growth firms relative to consumption good firms and value firms. In contrast, a neutral productivity shock affects all firms symmetrically. The presence of two shocks creates heterogeneity in expected returns and also explains why HML is a factor in the time series of returns, or equivalently why value and growth stocks move together.

The implications of the model are largely supported by the data. Using the restrictions of the model, I employ a portfolio of investment minus consumption firms (IMC) to proxy for investment-specific technological change. The IMC portfolio predicts an increase in investment, a fall in the quality-adjusted price of new equipment and a short run fall in consumption. I find that, controlling for their exposure to the market portfolio, firms with high correlation with the IMC portfolio have lower average returns. Additionally, the IMC portfolio improves upon the ability of the (C)CAPM to price the cross-section of expected returns of portfolios of stocks sorted by book to market.

More than simply adding an additional source of systematic risk, a model with investment-specific technological change offers new insights about the relation of stock returns and macroeconomic sources of risk. Similar to Gomes, Kogan and Yogo (2006), my model recognizes an aspect of firm heterogeneity that is not directly related to accounting valuation ratios and implies that firms respond differently to some macroeconomic shocks. My model contributes to the macroeconomic literature by illustrating how the cross-section of stock returns can help identify different macroeconomic shocks. My model also contributes to the finance literature by showing that a source of systematic risk previously considered in the macroeconomic literature can have important implications about the cross-section of risk premia and can help explain some of the stylized facts regarding value and growth firms.
1.6 Tables and figures

Table 1.1: IMC Portfolio: Composition

<table>
<thead>
<tr>
<th></th>
<th>Consumption Portfolio</th>
<th>Investment portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean min max</td>
<td>mean min max</td>
</tr>
<tr>
<td># of firms</td>
<td>2467 153 5858</td>
<td>832 44 2055</td>
</tr>
<tr>
<td>ME (USb)</td>
<td>2140.5 33.3 10951.5</td>
<td>529.7 2.4 4903.4</td>
</tr>
<tr>
<td>Book-to-Market Equity</td>
<td>0.62 0.32 1.14</td>
<td>0.54 0.12 1.06</td>
</tr>
<tr>
<td>Debt-to-Asset</td>
<td>0.18 0.10 0.24</td>
<td>0.16 0.05 0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Consumption Portfolio</th>
<th>Investment portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean min max</td>
<td>mean min max</td>
</tr>
<tr>
<td># of firms</td>
<td>1642 102 4265</td>
<td>364 23 841</td>
</tr>
<tr>
<td>ME (USb)</td>
<td>1709.9 25.3 9070.2</td>
<td>167.5 1.9 1019.1</td>
</tr>
<tr>
<td>Book-to-Market Equity</td>
<td>0.63 0.32 1.09</td>
<td>0.61 0.21 1.10</td>
</tr>
<tr>
<td>Debt-to-Asset</td>
<td>0.17 0.11 0.26</td>
<td>0.17 0.05 0.23</td>
</tr>
</tbody>
</table>

Table 1.2: IMC Portfolio: Correlations

<table>
<thead>
<tr>
<th>1951:2005</th>
<th>IMC</th>
<th>IMCX</th>
<th>IMCGKY</th>
<th>MKT</th>
<th>SMB</th>
<th>HML</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMC</td>
<td>82.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMCX</td>
<td></td>
<td>81.3%</td>
<td>73.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMCGKY</td>
<td>81.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MKT</td>
<td>42.8%</td>
<td></td>
<td>39.2%</td>
<td>36.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td>54.3%</td>
<td></td>
<td>54.0%</td>
<td>49.9%</td>
<td>25.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>-56.1%</td>
<td></td>
<td>-34.3%</td>
<td>-37.5%</td>
<td>-34.5%</td>
<td>-25.0%</td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>1.4%</td>
<td></td>
<td>-4.8%</td>
<td>0.2%</td>
<td>-5.7%</td>
<td>-1.4%</td>
<td>-13.0%</td>
</tr>
</tbody>
</table>

Table 1.1 reports composition details for the following two portfolios. IMC is the investment minus consumption portfolio constructed using the NIPA tables, where industries are classified as investment or consumption based on which sector they contribute the most. IMCX is the investment minus consumption portfolio constructed by excluding common industries. Data on book equity, long-term debt and total assets are from COMPUSTAT. Table 1.2 reports correlations with the above portfolios with the excess returns on the CRSP value-weighted index, the Fama-French (1993) factors, Carhart’s momentum factor and IMCGKY. IMCGKY is the investment minus consumption constructed using the data provided by Gomes, Kogan and Yogo (2006). I form IMCGKY by subtracting one half times the services portfolio and one half times the non-durables portfolio from the investment portfolio. I report results for the whole sample (1951-2005).
Table 1.3: Summary Statistics

<table>
<thead>
<tr>
<th>Period</th>
<th>IMC</th>
<th>IMCX</th>
<th>IMCGKY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E(R) α cor(R, R_{HML})</td>
<td>E(R) α cor(R, R_{HML})</td>
<td>E(R) α cor(R, R_{HML})</td>
</tr>
<tr>
<td>1951-1960</td>
<td>-0.03 -1.26 0.14</td>
<td>-2.46 -3.62 0.30</td>
<td>1.80 -0.71 0.29</td>
</tr>
<tr>
<td>1961-1970</td>
<td>-0.11 -1.26 -0.34</td>
<td>0.06 -1.14 -0.18</td>
<td>-0.38 -1.30 -0.33</td>
</tr>
<tr>
<td>1971-1980</td>
<td>3.17 2.02 -0.27</td>
<td>4.72 3.65 -0.18</td>
<td>0.80 0.22 -0.00</td>
</tr>
<tr>
<td>1981-1990</td>
<td>-8.63 -9.98 -0.67</td>
<td>-11.93 -13.20 -0.51</td>
<td>-10.70 -11.28 -0.37</td>
</tr>
<tr>
<td>1991-2000</td>
<td>4.76 -2.61 -0.78</td>
<td>-2.62 -7.88 -0.51</td>
<td>-1.05 -6.17 -0.55</td>
</tr>
<tr>
<td>2001-2005</td>
<td>-3.65 -4.99 -0.77</td>
<td>3.60 2.51 -0.46</td>
<td>-0.51 -1.11 -0.71</td>
</tr>
<tr>
<td>1951-2005</td>
<td>-0.49 -3.11 -0.56</td>
<td>-1.90 -4.12 -0.34</td>
<td>-1.80 -3.60 -0.37</td>
</tr>
</tbody>
</table>

Table reports average annualized excess returns, CAPM alphas and correlations with HML for the following three portfolios. IMC is the investment minus consumption portfolio constructed using the NIPA tables, where industries are classified as investment or consumption based on which sector they contribute the most. IMCX is the investment minus consumption portfolio constructed by excluding common industries. Sample includes data from 1951 to 2005. IMCGKY is the investment minus consumption constructed using the data of Gomes, Kogan and Yogo (2006). I form IMCGKY by subtracting one half times the services portfolio and one half times the non-durables portfolio from the investment portfolio. I report results for the whole sample and for each decade.
For every stock in CRSP with more than 5 years of data I estimate the regression

\[ r_{it} = a + \beta_i f_t + u_t \]

where \( r_{it} \) are log returns of the individual stock and \( f_t \) are log returns on factors (log excess returns on the market portfolio, HML or IMC). I compute the \( R^2 \) for each regression, also with a degrees of freedom adjustment (\( \bar{R}^2 \)), and report the equal-weighted average across stocks (EW) along with the average weighted by the individual stock variance (VW). The sample includes data from 1951 to 2005 and 12,953 stocks.

<table>
<thead>
<tr>
<th>Factors</th>
<th>EW ( R^2 )</th>
<th>( \bar{R}^2 )</th>
<th>VW ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>13.7%</td>
<td>13.7%</td>
<td>12.0%</td>
</tr>
<tr>
<td>MKT + IMC</td>
<td>15.9%</td>
<td>15.3%</td>
<td>14.8%</td>
</tr>
<tr>
<td>MKT + HML</td>
<td>16.0%</td>
<td>15.3%</td>
<td>14.6%</td>
</tr>
<tr>
<td>MKT + HML + IMC</td>
<td>17.6%</td>
<td>16.3%</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

The sample includes data from 1951 to 2005 and 12,953 stocks.
The top figure shows the ratio of market values of investment over consumption industries and the ratio of private fixed nonresidential investment over personal consumption expenditures from the NIPA tables (dotted line). The solid line uses the baseline classification, where industries are assigned to the sector the contribute the most, whereas the dotted line represents the ratio of market values when I drop common industries. The bottom figure shows returns for the IMC portfolio, with the solid line classifying industries according to the sector they contribute most, and the dotted line drops common industries.
Table 1.5: Summary Statistics: 10 portfolios sorted on IMC beta

<table>
<thead>
<tr>
<th>IMC Beta</th>
<th>Lo</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Return (%)</td>
<td>6.34</td>
<td>7.87</td>
<td>7.20</td>
<td>6.78</td>
<td>6.80</td>
<td>6.21</td>
<td>5.75</td>
<td>5.86</td>
<td>5.30</td>
<td>5.85</td>
</tr>
<tr>
<td>σ (%)</td>
<td>13.05</td>
<td>15.14</td>
<td>14.71</td>
<td>16.05</td>
<td>16.78</td>
<td>16.80</td>
<td>18.74</td>
<td>20.30</td>
<td>22.71</td>
<td>25.68</td>
</tr>
<tr>
<td>β_{MKT}</td>
<td>0.66</td>
<td>0.91</td>
<td>0.88</td>
<td>0.97</td>
<td>1.02</td>
<td>1.00</td>
<td>1.13</td>
<td>1.22</td>
<td>1.35</td>
<td>1.44</td>
</tr>
<tr>
<td>α(%)</td>
<td>2.60</td>
<td>2.68</td>
<td>2.17</td>
<td>1.22</td>
<td>0.98</td>
<td>0.49</td>
<td>-0.73</td>
<td>-1.11</td>
<td>-2.38</td>
<td>-2.40</td>
</tr>
</tbody>
</table>

| β_{MKT}  | 0.81 | 0.99 | 0.96 | 1.00 | 1.03 | 1.05 | 1.10 | 1.16 | 1.22 | 1.28 |
| β_{SMB}  | -0.35 | -0.23 | -0.19 | -0.13 | -0.07 | 0.07 | 0.09 | 0.24 | 0.28 | 0.51 |
| β_{HML}  | 0.32 | 0.11 | 0.16 | 0.00 | -0.01 | 0.23 | -0.05 | -0.04 | -0.24 | -0.20 |
| α(%)     | 0.99 | 2.32 | 1.41 | 1.40 | 1.15 | -1.14 | -0.55 | -1.19 | -1.24 | -1.88 |

Table reports summary statistics of the simple monthly returns of 10 portfolios created by sorting stocks on their covariance with IMC. The pre-sorting covariances are estimated using weekly returns and 5 year windows. Portfolios are rebalanced every 5 years on June. Panel A reports mean excess returns over the 30-day T-bill rate, the standard deviation of returns and their beta and alpha with respect to the CAPM. Panel B reports beta and alpha for the Fama-French three factor model. Panel C reports beta and alpha for the model including the market portfolio and the IMC portfolio.
Table 1.6: *Cross-Sectional Tests: 24 portfolios sorted on Industry and IMC covariance*

<table>
<thead>
<tr>
<th>Factor Price</th>
<th>MAPE(%)</th>
<th>SSQE(%)</th>
<th>J-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>1.11</td>
<td>2.93</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(2.37)</td>
<td>(1.01)</td>
</tr>
<tr>
<td>SMB</td>
<td>-5.66</td>
<td>-1.58</td>
<td>-5.66</td>
</tr>
<tr>
<td></td>
<td>(-2.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>2.76</td>
<td>32.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.01)</td>
<td>(0.70)</td>
<td></td>
</tr>
<tr>
<td>$C_{ND}$</td>
<td>16.65</td>
<td>32.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(1.41)</td>
<td></td>
</tr>
<tr>
<td>IMC</td>
<td>-3.74</td>
<td>-4.39</td>
<td>-4.51</td>
</tr>
<tr>
<td></td>
<td>(-2.71)</td>
<td>(-2.51)</td>
<td>(-2.27)</td>
</tr>
<tr>
<td>MAPE(%)</td>
<td>1.72</td>
<td>0.98</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.104)</td>
<td></td>
</tr>
<tr>
<td>SSQE(%)</td>
<td>1.23</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.343)</td>
<td>(0.330)</td>
<td></td>
</tr>
<tr>
<td>J-test</td>
<td>1.17</td>
<td>0.47</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>(0.330)</td>
<td>(0.521)</td>
<td></td>
</tr>
</tbody>
</table>

Table reports the estimation results from cross-sectional tests. I estimate the parameters of the pricing kernel,

$$\pi = a - b F.$$ 

The set of test assets includes simple monthly returns of 24 portfolios created by sorting stocks first on 8 industries and then on their covariance with IMC. The pre-sorting covariances are estimated using weekly returns and 5 year windows. Portfolios are rebalanced every 5 years on June. The factors considered are the market portfolio, Fama and French’s SMB and HML factors, the IMC portfolio, non-durable consumption growth ($C_{ND}$). The sample includes monthly data from 1959 to 2005. Estimation is by two-step generalized method of moments (GMM). HAC standard errors are computed by the Newey-West estimator with 3 lags of returns. I report second stage estimates of the factor premia (b), along with t-statistics, annualized mean absolute pricing errors (MAPE) and the sum of squared pricing errors (SSQE) for the first stage. Furthermore, I report the J-test that all the pricing errors are zero along with it’s p-value in parenthesis.
Table 1.7: Cross-Sectional Tests: 25 portfolios sorted on IMC and MKT beta

<table>
<thead>
<tr>
<th>Factor Price</th>
<th>MAPE(%)</th>
<th>SSQE(%)</th>
<th>J-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>2.45</td>
<td>3.01</td>
<td>3.69</td>
</tr>
<tr>
<td></td>
<td>(2.30)</td>
<td>(2.37)</td>
<td>(2.99)</td>
</tr>
<tr>
<td>SMB</td>
<td>0.26</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>2.98</td>
<td>-1.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_{ND}$</td>
<td>23.45</td>
<td>33.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td>(2.80)</td>
<td></td>
</tr>
<tr>
<td>IMC</td>
<td>-3.10</td>
<td>-1.61</td>
<td>-3.67</td>
</tr>
<tr>
<td></td>
<td>(-2.01)</td>
<td>(-2.10)</td>
<td>(-1.76)</td>
</tr>
<tr>
<td>MAPE(%)</td>
<td>2.22</td>
<td>2.44</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(0.354)</td>
<td>(0.453)</td>
<td>(0.339)</td>
</tr>
<tr>
<td>SSQE(%)</td>
<td>1.58</td>
<td>1.80</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>(0.453)</td>
<td>(0.339)</td>
<td>(0.572)</td>
</tr>
<tr>
<td>J-test</td>
<td>26.0</td>
<td>24.1</td>
<td>24.2</td>
</tr>
<tr>
<td></td>
<td>(0.354)</td>
<td>(0.453)</td>
<td>(0.339)</td>
</tr>
</tbody>
</table>

Table reports the estimation results from cross-sectional tests. I estimate the parameters of the pricing kernel,

$$\pi = a - b F.$$  

The set of test assets includes simple monthly returns of 25 portfolios created by sorting stocks first on their market beta and then on their beta with IMC. The pre-sorting covariances are estimated using weekly returns and 5 year windows. Portfolios are rebalanced every 5 years on June. The factors considered are the market portfolio, Fama and French’s SMB and HML factors, the IMC portfolio, non-durable consumption growth ($C_{ND}$). The sample includes monthly data from 1959 to 2005. Estimation is by two-step generalized method of moments (GMM). HAC standard errors are computed by the Newey-West estimator with 3 lags of returns. I report second stage estimates of the factor premia (b), along with t-statistics, annualized mean absolute pricing errors (MAPE) and the sum of squared pricing errors (SSQE) for the first stage. Furthermore, I report the J-test that all the pricing errors are zero along with it’s p-value in parenthesis.
Table 1.8: Cross-Sectional Tests: 25 portfolios sorted on ME and IMC covariance

<table>
<thead>
<tr>
<th>Factor Price</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>2.26</td>
<td>3.79</td>
<td>3.48</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(2.80)</td>
<td>(2.87)</td>
<td>(2.88)</td>
</tr>
<tr>
<td>SMB</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>7.40</td>
<td></td>
<td>3.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td></td>
<td>(0.82)</td>
<td></td>
</tr>
<tr>
<td>$C_{ND}$</td>
<td>26.89</td>
<td></td>
<td>155.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td></td>
<td>(3.17)</td>
<td></td>
</tr>
<tr>
<td>IMC</td>
<td>-3.50</td>
<td>-4.28</td>
<td>-2.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.29)</td>
<td>(-2.25)</td>
<td>(-1.02)</td>
<td></td>
</tr>
<tr>
<td>MAPE(%)</td>
<td>2.06</td>
<td>2.26</td>
<td>0.91</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>(1.24)</td>
<td>(1.25)</td>
<td>(0.53)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>SSQE(%)</td>
<td>1.48</td>
<td>1.70</td>
<td>0.28</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
<td>(0.29)</td>
<td>(0.59)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>J-test</td>
<td>27.5</td>
<td>27.1</td>
<td>19.8</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.29)</td>
<td>(0.59)</td>
<td>(0.80)</td>
</tr>
</tbody>
</table>

Table reports the estimation results from cross-sectional tests. I estimate the parameters of the pricing kernel,

$$\pi = a - bF.$$  

The set of test assets includes simple monthly returns of 25 portfolios created by sorting stocks first on their market capitalization based on NYSE breakpoints and then on their covariance with IMC. The pre-sorting covariances are estimated using weekly returns and 5 year windows. Portfolios are rebalanced every 5 years on June. The factors considered are the market portfolio, Fama and French's SMB and HML factors, the IMC portfolio, non-durable consumption growth ($C_{ND}$). The sample includes monthly data from 1959 to 2005. Estimation is by two-step generalized method of moments (GMM). HAC standard errors are computed by the Newey-West estimator with 3 lags of returns. I report second stage estimates of the factor premia ($b$), along with t-statistics, annualized mean absolute pricing errors (MAPE) and the sum of squared pricing errors (SSQE) for the first stage. Furthermore, I report the J-test that all the pricing errors are zero along with it's p-value in parenthesis.
Table 1.9: Cross-Sectional Tests: 25 portfolios sorted on BM and ME

<table>
<thead>
<tr>
<th>Factor Price</th>
<th>Market</th>
<th>SMB</th>
<th>HML</th>
<th>CND</th>
<th>IMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.61</td>
<td>3.11</td>
<td>9.12</td>
<td>90.79</td>
<td>9.12</td>
</tr>
<tr>
<td></td>
<td>(2.51)</td>
<td>(2.13)</td>
<td>(5.18)</td>
<td>(4.21)</td>
<td>(5.18)</td>
</tr>
<tr>
<td></td>
<td>4.09</td>
<td>7.79</td>
<td>12.0</td>
<td>-10.83</td>
<td>-4.04</td>
</tr>
<tr>
<td></td>
<td>(3.52)</td>
<td>(3.63)</td>
<td>(3.00)</td>
<td>(-3.99)</td>
<td>(-2.28)</td>
</tr>
<tr>
<td></td>
<td>4.90</td>
<td>3.54</td>
<td>5.35</td>
<td>-5.35</td>
<td>-5.35</td>
</tr>
<tr>
<td></td>
<td>(3.29)</td>
<td>(1.42)</td>
<td>(1.42)</td>
<td>(-2.69)</td>
<td>(-2.69)</td>
</tr>
<tr>
<td></td>
<td>3.90</td>
<td>1.30</td>
<td>0.58</td>
<td>-0.80</td>
<td>-0.80</td>
</tr>
<tr>
<td></td>
<td>(3.08)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(-0.19)</td>
<td>(-0.19)</td>
</tr>
<tr>
<td></td>
<td>4.15</td>
<td>2.53</td>
<td>1.30</td>
<td>1.16</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>(3.08)</td>
<td>(2.28)</td>
<td>(2.28)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

Table reports the estimation results from cross-sectional tests. I estimate the parameters of the pricing kernel,

\[ \pi = a - b F. \]

The set of test assets includes simple monthly returns of the 25 portfolios created by sorting stocks on first market capitalization and then on book to market using NYSE quintiles. The data come from Kenneth French’s website. The factors considered are the market portfolio, Fama and French’s SMB and HML factors, the IMC portfolio, non-durable consumption growth \( C_{ND} \). The sample includes monthly data from 1959 to 2005. Estimation is by two-step generalized method of moments (GMM). HAC standard errors are computed by the Newey-West estimator with 3 lags of returns. I report second stage estimates of the factor premia \( b \), along with t-statistics, annualized mean absolute pricing errors (MAPE) and the sum of squared pricing errors (SSQE) for the first stage. Furthermore, I report the J-test that all the pricing errors are zero along with it’s p-value in parenthesis.
Table 1.10: Cross-Sectional Tests: 25 portfolios sorted on BM and IMC covariance

<table>
<thead>
<tr>
<th>Factor Price</th>
<th>Market</th>
<th>SMB</th>
<th>HML</th>
<th>CND</th>
<th>IMC</th>
<th>MAPE(%)</th>
<th>SSQE(%)</th>
<th>J-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.80</td>
<td>-0.34</td>
<td>8.13</td>
<td>4.70</td>
<td>4.02</td>
<td>4.86</td>
<td>(2.59)</td>
<td>87.36</td>
</tr>
<tr>
<td></td>
<td>(2.59)</td>
<td>(-0.15)</td>
<td>(3.45)</td>
<td>(3.64)</td>
<td>(3.23)</td>
<td>(3.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CND</td>
<td>87.36</td>
<td>7.68</td>
<td>(3.00)</td>
<td>8.13</td>
<td>3.45</td>
<td>-0.34</td>
<td>(-2.28)</td>
<td>(-0.58)</td>
</tr>
<tr>
<td></td>
<td>(3.00)</td>
<td>(3.12)</td>
<td>(2.80)</td>
<td>(3.45)</td>
<td>(-2.07)</td>
<td>(-0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE(%)</td>
<td>2.42</td>
<td>1.20</td>
<td>1.65</td>
<td>1.43</td>
<td>1.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSQE(%)</td>
<td>2.22</td>
<td>0.57</td>
<td>1.08</td>
<td>0.87</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J-test</td>
<td>44.8</td>
<td>34.0</td>
<td>40.5</td>
<td>22.8</td>
<td>34.2</td>
<td>(0.006)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.049)</td>
<td>(0.472)</td>
<td>(0.472)</td>
<td>(0.472)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

Table reports the estimation results from cross-sectional tests. I estimate the parameters of the pricing kernel,

$$\pi = a - b F.$$

The set of test assets includes simple monthly returns of 25 portfolios created by sorting stocks first on their book to market ratio based on NYSE breakpoints and then on their covariance with IMC. The pre-sorting covariances are estimated using weekly returns and 5 year windows. Portfolios are rebalanced every 5 years on June. The factors considered are the market portfolio, Fama and French’s SMB and HML factors, the IMC portfolio, non-durable consumption growth ($C_{ND}$). The sample includes monthly data from 1959 to 2005. Estimation is by two-step generalized method of moments (GMM). HAC standard errors are computed by the Newey-West estimator with 3 lags of returns. I report second stage estimates of the factor premia (b), along with t-statistics, annualized mean absolute pricing errors (MAPE) and the sum of squared pricing errors (SSQE) for the first stage. Furthermore, I report the J-test that all the pricing errors are zero along with it’s p-value in parenthesis.
Figure plots the CAPM pricing errors of the test portfolios considered in tables 5 through 10 versus their covariances with the IMC portfolio. CAPM Pricing errors are the first stage pricing errors from the cross-sectional GMM tests. The constructions of these portfolios is standard and is described in the Appendix.
<table>
<thead>
<tr>
<th></th>
<th>$cov(R_i, R_{MKT})$</th>
<th>$cov(R_i, R_{IMC})$</th>
<th>$\ln ME$</th>
<th>$\ln BM$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: 1951 - 2005</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.626</td>
<td>-0.116</td>
<td>0.265</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(-3.28)</td>
<td>(4.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.086</td>
<td>-3.694</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td>(-2.16)</td>
<td>(-3.37)</td>
<td>(4.45)</td>
</tr>
<tr>
<td><strong>Panel B: 1951 - 1981</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.149</td>
<td>-0.128</td>
<td>0.126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
<td>(-2.75)</td>
<td>(1.46)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.122</td>
<td>-1.916</td>
<td>-0.129</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(-1.26)</td>
<td>(-2.80)</td>
<td>(1.39)</td>
</tr>
<tr>
<td><strong>Panel C: 1982 - 2005</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td>-0.103</td>
<td>0.410</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(-1.93)</td>
<td>(5.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.090</td>
<td>-5.546</td>
<td>-0.105</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(-2.59)</td>
<td>(-2.00)</td>
<td>(5.19)</td>
</tr>
</tbody>
</table>

Table 1.11: IMC and Expected Returns - Entire Cross-section of stocks

Table reports results from Fama and MacBeth (1973) cross-sectional regressions of simple monthly returns of all NYSE, AMEX, and Nasdaq stocks on covariances and characteristics. Covariances are estimated using a procedure similar to Fama and French (1992). Specifically, for each individual stock I estimate the covariance of its returns with IMC using 5 years of weekly log excess returns. At the end of a five year period, stocks are then sorted into 100 pre-ranking covariance centiles. I then compute the equal-weighted monthly log excess returns on these 100 portfolios over the next 5 years. This procedure is repeated every 5 years, forming a time-series of returns on these 100 portfolios. I then reestimate covariances for the portfolios formed from the pre-ranking sorts using 5 years of monthly data to obtain post-ranking covariances. The post-ranking covariance estimate for a given portfolio is then assigned to each stock in the portfolio. Portfolio assignments are updated every 5 years. Every month the cross-section of stock returns in excess of risk free rate is then regressed on a constant (not reported), the covariance with the excess return on the CRSP value-weighted index, the covariance with the return on the IMC portfolio, the log of market capitalization (ME) and the log of Book to Market (BM). I report the time series average of the regression coefficients along with t-statistics using Fama-McBeth errors. Panel A reports results for the full sample, whereas Panels B and C report results for the first and second subperiod.
Figure 1-3: Cumulative response on IMC: Investment

![Graphs of investment and relative prices](image)

(a) Investment Information processing equipment and software
(b) Relative price: Information processing equipment and software
(c) Investment in Industrial Equipment
(d) Investment in Non-Residential structures

Figure plots the coefficients $\beta_k$ in the regression of:

$$x_{t+k} - x_t = a + \beta_k IMC_t + \gamma_k MKT_t + \epsilon_{tk}$$

along with two HAC standard error bands. I compute HAC standard errors using Newey-West, with the truncation lag equal to the length of the overlap plus two quarters. The sample includes quarterly data in the 1951:2005 period. The quantity and price indices are from the NIPA tables in the BEA website. The price index for Information processing equipment and software is relative to the GDP deflator.
Figure 1-4: Cumulative response on IMC: Consumption and output

(a) Personal consumption expenditures excluding food and energy
(b) Personal consumption expenditures: Durable Goods
(c) Residential Investment
(d) Gross Domestic Product excluding government expenditures

Figure plots the coefficients $\beta_k$ in the regression of:

$$x_{t+k} - x_t = a + \beta_k IMC_t + \gamma_k MKT_t + \epsilon_{t+k}$$

along with two HAC standard error bands. I compute HAC standard errors using Newey-West, with the truncation lag equal to the length of the overlap plus two quarters. The sample includes quarterly data in the 1951:2005 period. The quantity indices are from the NIPA tables in the BEA website.
Table 1.12: Real rates and IMC(X)

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Level / Diff/Ret</th>
<th>Include common industries?</th>
<th>$\hat{\beta} \times 100$</th>
<th>t</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Y</td>
<td>L</td>
<td>Y</td>
<td>3.45</td>
<td>5.3</td>
<td>28.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>3.69</td>
<td>3.7</td>
<td>16.0%</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>Y</td>
<td>0.44</td>
<td>2.5</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>0.92</td>
<td>2.3</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Y</td>
<td>0.86</td>
<td>2.1</td>
<td>0.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>0.65</td>
<td>1.7</td>
<td>0.6%</td>
</tr>
<tr>
<td>1 Y</td>
<td>L</td>
<td>Y</td>
<td>1.80</td>
<td>2.1</td>
<td>9.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>1.57</td>
<td>1.2</td>
<td>3.5%</td>
</tr>
<tr>
<td></td>
<td>Δ</td>
<td>Y</td>
<td>0.76</td>
<td>1.3</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>0.04</td>
<td>0.1</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Y</td>
<td>-0.00</td>
<td>-0.5</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>-0.81</td>
<td>-1.0</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Table reports results for a regressions of real rates on ratio of market capitalizations. I estimate the equations

$$r_{t,t+T}^n - \pi_{t,t+T} = \alpha + \beta \hat{\omega}_t + u_t$$  \hspace{1cm} (L)

$$\Delta r_{t,t+T}^n - \Delta \pi_{t,t+T} = \alpha + \beta \Delta \hat{\omega}_t + u_t$$  \hspace{1cm} (Δ)

$$\Delta r_{t,t+T}^n - \Delta \pi_{t,t+T} = \alpha + \beta R_{IMC,t} + u_t$$  \hspace{1cm} (R)

where $r_{t,t+T}^n$ is the nominal yield on a Treasury bond with maturity T, $\pi_{t,t+T}$ is realized inflation based on the CPI-U index between time t and t+T, $\hat{\omega}$ is the log ratio of market capitalization of the investment versus the consumption sector. Sample includes monthly data from 1953 to 2005. I report results using both classifications schemes. The first, IMC, assigns industries to investment or consumption based on which sector they contribute the most, whereas the second, IMCX, excludes common industries. Because both variables are close to non-stationary, I estimate the regression in levels (L), first differences (Δ). Additionally, I estimate the equation in first differences replacing $\Delta \hat{\omega}$ with returns on the IMC(X) portfolio (R). I report OLS estimates along with t-statistics calculated using Newey-West errors.
Table 1.13: IMC and Investment: Firm Level Evidence

<table>
<thead>
<tr>
<th>$I_{i,t+1}/K_{i,t}$</th>
<th>$q_{i,t}$</th>
<th>$\beta_{i,t}^{MC}$</th>
<th>$R_{i,t}^{MC}$</th>
<th>$\beta_{i,t}^{MC} \times R_{i,t}^{MC}$</th>
<th>$q_{i,t} \times R_{i,t}^{MC}$</th>
<th>FFE</th>
<th>TFE</th>
<th>IND x T FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPEX$<em>{i,t+1}/PPE</em>{i,t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.065***</td>
<td>0.006*</td>
<td>0.066***</td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>0.065***</td>
<td>0.008**</td>
<td>0.034***</td>
<td>0.064***</td>
<td>-0.002</td>
<td></td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>0.076***</td>
<td>0.002</td>
<td>0.075***</td>
<td>0.005</td>
<td></td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>0.055***</td>
<td>0.014***</td>
<td>0.064*</td>
<td>0.010</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>0.075***</td>
<td>0.007</td>
<td>0.140***</td>
<td>0.006</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>CAPEX$<em>{i,t+1}/A</em>{i,t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.023***</td>
<td>0.002***</td>
<td>0.028***</td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>0.023***</td>
<td>0.002***</td>
<td>0.022***</td>
<td>0.013***</td>
<td>0.001</td>
<td></td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>0.027***</td>
<td>0.000</td>
<td>0.007***</td>
<td>0.006***</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>0.017***</td>
<td>0.002***</td>
<td>0.010**</td>
<td>0.009**</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>0.025***</td>
<td>0.001</td>
<td>0.012***</td>
<td>0.006***</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

The table shows coefficients of a regression of firm level investment ($I_{i,t+1}/K_{i,t}$) on beginning of period log Market Equity to Book Equity ratio ($q_{i,t}$), firm beta with respect to IMC ($\beta_{i,t}$), returns on the IMC portfolio over the past year ($R_{i,t}^{MC}$). Standard errors are clustered at the firm level. The column labeled FFE indicates if firm fixed effects are included, the column TFE indicates if time fixed effects are used and the column IND x T FE indicates if industry-year fixed effects are included. Investment is defined as either CAPEX (COMPSTAT data item 128) over Property, Plant and Equipment (COMPSTAT data item 8) or CAPEX over Total Assets (data item 6). Firm level betas are estimated using weekly stock returns using annual, non-overlapping windows. Weekly stock returns are constructed by compounding daily returns. Industries are defined by 2-digit SIC codes. The sample includes 5131 firms during the 1951-2005 period. I exclude firms that have less than 10 years of data in COMPSTAT and firms with less than 50 weekly return observations in CRSP.

Statistical Significance level * 10%  ** 5%  *** 1%.
Chapter 2
Sources of systematic risk

This chapter is based on work with Igor Makarov.¹

Abstract
In this chapter we use heteroscedasticity of stock returns as an identification tool to isolate four robust factors in the U.S. industry returns. The first factor can be viewed as a proxy for economy wide demand shocks. The second factor is a portfolio of stocks producing investment goods minus stocks producing consumption goods (IMC). The third factor differentiates between cyclical vs. non-cyclical stocks. Finally, the fourth factor is consistent with a proxy for shocks to input good prices. The extracted factors are shown to be important in explaining the cross-section of expected returns. Unlike the CAPM or the Fama and French three factor model, they successfully price the cross-section of 48 industry portfolios and do a good job at explaining the 25 Fama and French size and book-to-market portfolios. The fourth (“input”) factor is found to be a robust predictor of the value-weighted market portfolio.

2.1 Introduction
Since stock represent a claim on the future output of the economy, stock returns should respond to news about the economy. This idea lies at the heart of finance and an extensive literature has aimed to resolve the apparent disconnect between stock price movements and economic fundamentals. Diversification arguments suggest to focus on pervasive or “systematic” factors. Many possible ”macro-factors” have been proposed, including technology, money, credit, oil, and others, but commonly used macroeconomic proxies for these shocks fail to explain persuasively fluctuations in financial markets. Moreover, they seem to have

¹The authors would like to thank Moto Yogo and Kenneth French for sharing with us their data and Alexei Onatski and Serena Ng for sharing with us their code. We are grateful to Vito Gala, Leonid Kogan, Alexei Onatski, Roberto Rigobon, and Stephen Ross for valuable discussions and comments. We thank participants at MIT Finance lunch and LBS Finance lunch for helpful comments.
a hard time accounting for a significant fraction of economic fluctuations (Cochrane 1994). One popular approach starting with Chen, Roll, and Ross (1986) is to look at innovations to macroeconomic variables which trace the state of the economy. This approach requires one to specify in advance what the economic factors are. Unfortunately, the body of empirical evidence again is mixed and inconclusive.

In this chapter we take a different stand and use stock returns themselves in order to understand the underlying forces behind stock price movements. We will be agnostic as to what the sources of systematic risk are, but rather let the data speak for itself. The idea to use a subset of stock returns to explain other asset returns is not new, but it is central to factor analysis. However, its application to asset pricing has been hindered by a major drawback. Under the standard assumptions of constant factor covariance matrix one cannot identify individual factors, but only their linear span. The factors can be arbitrary rotated: If \( f_1 \) and \( f_2 \) are factors, so are \( f_1 - f_2 \) and \( f_1 + f_2 \). This makes interpretation difficult and often leads to unstable factors, i.e. factors extracted from different sub-samples do not resemble each other.

At the same time, the world is far from homoscedastic. Many macroeconomic quantities, stock returns, and even consumption growth display time-varying volatility. While econometrics textbooks often treat heteroscedasticity as a problem which has to be dealt with to get the inference right, we take it as a blessing rather than a curse. Relaxing the standard assumptions of factor analysis, by allowing the covariance matrix of factors to change over time, results in exactly identified factors. That is, factors are not invariant to rotation. Identifications essentially follows from requiring the factors to be orthogonal at every point in time.\(^3\) The assumption of multiple heteroscedastic factors also has intuitive appeal. It implies that there are multiple sources of systematic risk, whose relative importance varies over time. Casual empiricism seems to support this view. In the late 90's everyone was thrilled by internet revolution and little attention was paid to oil prices. In contrast, nowadays almost every issue of Wall Street Journal discusses how stock prices react to changes in oil prices.

Accounting for heteroscedasticity improves significantly the stability of the factors. We demonstrate that analysis applied to different sub-samples identifies essentially the same factors, thus alleviating concern of sample dependence associated with homoscedastic version of factor analysis. At the same time, factor heteroscedasticity suggests a possible explanation for the observed instability of previous applications of the method. Inability to pinpoint the right factors from their linear span often results in arranging factors that are ranked according

---

\(^2\) An incomplete list includes Schwert (1989), Bansal and Yaron (2004).

\(^3\) In the homoscedastic case the constraint is degenerate resulting in under-identification. To better see it consider a simple example. Suppose we have two orthogonal factors \( f_{1,t} \) and \( f_{2,t} \) whose variances at time \( t \) equal \( \sigma_{1,t}^2 \) and \( \sigma_{2,t}^2 \) respectively. Without loss of generality we assume that each factor has a unit unconditional variance. If \( \tilde{f}_{1,t} = (f_{1,t} + f_{2,t})/2 \) and \( \tilde{f}_{2,t} = (f_{1,t} - f_{2,t})/2 \) is a linear transformation of the original factors then \( E_t(\tilde{f}_{1,t}\tilde{f}_{2,t}) = (\sigma_{1,t}^2 - \sigma_{2,t}^2)/4 \). So the transformed factors are orthogonal at time \( t \) if and only if \( \sigma_{1,t}^2 = \sigma_{2,t}^2 \). This is true for all \( t \) in the homoscedastic case, and therefore, factors are not identified in this case. In contrast, in the heteroscedastic case, factors are identified, provided that there is a \( t \) when \( \sigma_{1,t}^2 \neq \sigma_{2,t}^2 \).
to their ability to explain the amount of variation in stock returns. But time-varying factor volatilities make this ranking very time dependent and noisy.

Using identification through heteroscedasticity we isolate four robust factors from industry returns. The first factor has similar loadings on all industries and explains the highest proportion of variation in realized returns. Unsurprisingly, it is highly correlated with the market portfolio. The second factor is highly correlated (87%) with a portfolio of stocks producing investment goods minus stocks producing consumption goods (IMC) and likely captures shocks to the relative price of capital to consumption goods. The third factor differentiates between cyclical vs. non-cyclical industries, such as Auto manufacturers versus Service providers, and likely captures information related to the business cycle. Finally, the fourth factor loads positively on industries producing commodities such as oil, gas, metals and negatively with the rest of the market. Thus, it likely captures changes in the relative price of input goods.

Additionally, the extracted factors are shown to be important in explaining the cross-section of expected returns. Unlike the CAPM or the Fama and French three factor model, they successfully price the cross-section of 48 industry portfolios. In addition, they represent a significant improvement over the CAPM in pricing 25 Fama and French size and book-to-market portfolios. Finally, we sort all stocks into portfolios based on their loadings on our four factors. In all cases except the first factor, we show that this results in an almost monotone pattern for abnormal returns.

Moreover, the fourth (“input”) factor is found to be a good predictor of the value-weighted market portfolio both in-sample and out-of-sample. It also contains information in addition to well known predictors such as the short rate, term spread and dividend yield. This is consistent with the results of Hong Torous and Valkanov who find that 12 out of 38 industries lead the market. However, we find that their result is entirely driven by our fourth factor. Once this factor is included in the predictive regression, no industry is a statistically significant predictor of the market. While Hong, Torous, and Valkanov (2006) interpret their findings as an indication of investors’ bounded rationality, our results point towards a more rational explanation. A fall in discount rates could induce an increase in investment and thus in the relative price of input goods.

Furthermore, our extracted factors contain information about macroeconomic variables. The first factor predicts quantities and prices with an opposite sign, and thus likely captures a supply side shock. The second and fourth factor, which capture changes in the relative price of capital and input goods, predict inflation. Finally the third factor predicts industrial production, is correlated with consumers expectations about business conditions and predicts a fall in the default spread.

The existence of multiple heteroscedastic factors provides a simple explanation for the observed volatile behavior of market betas over time. Intuitively, if the market portfolio is a sum of heteroscedastic shocks, then market beta of a stock with high loading on the factor whose volatility goes up should also increase. In fact, N-factor structure in returns

---

5See e.g. Harvey (1989), Fama and French (1997), Lewellen and Nagel (2006), and Andersen et. al. (2006) among others and references therein.
implies \((N-1)\)-factor structure in market betas. Consistent with the model we show that industry betas indeed comove and fluctuations in conditional betas match well their empirical counterparts.

The rest of the paper is organized as follows. In the next section we briefly review related literature. Section 3 lays out econometric framework, explains how factors are estimated and identified. Section 4 gives details of empirical implementation and describes the data. In section 5 we present our findings which include identification of extracted factors, results of asset pricing tests, and predictive regressions. Section 6 examine implication of the model for the dynamics of market betas. Section 7 concludes.

2.2 Related literature

A large body of the literature is interested in linking macroeconomic variables to the stock returns. Chen, Roll, and Ross (1986) show that default and term premia are priced risk factors, and that industrial production growth is a strong candidate. However, Shanken and Weinstein (1990) show that their conclusions depend on the specific method used to form test portfolios. Campbell (1996) presents evidence that revisions in forecast of future labor income growth is a risk factor that helps price the cross-section of stock returns. Vassalou (2003) documents that news related to future GDP growth are related to Fama and French HML and SMB factors. Petkova (2006) relates the HML and SMB factors to innovations in the short-term T-bill, term spread, aggregate dividend yield, and default spread. The above papers concentrate mostly on explaining the cross-section of expected returns, whereas in our paper we are interested in the shocks themselves. Another approach is pursued in Lamont (2000) who uses bond and industry returns to construct factor mimicking portfolios for expectations about future economic variables. He concludes that created portfolios can be useful in forecasting macroeconomic variables. Connor and Korajczyk (1991) project Chen, Roll, and Ross factors on factors extracted from stock returns to study the performance of mutual funds.

Inspired by the arbitrage pricing theory developed in Ross (1976) multi-factor models become popular in finance starting from late 80's. On the methodology side, we are in the domain of classical factor analysis when \(N\) is relatively small and fixed but \(T \to \infty\) (Lawley and Maxwell 1971). Connor and Korajczuk (1988) developed asymptotic principal component analysis to estimate approximate factor models when \(T\) is fixed and \(N \to \infty\). Jones (2001) extends their analysis to allow for time series heteroscedasticity in the factor model residuals. Stock and Watson (2002) consider the case when \(N, T \to \infty\) jointly. Ludvigson and Ng (2006) apply the later approach to forecast excess return and volatility of the market portfolio.

Our identification of factors is based on the results in Sentana and Fiorentini (2001). They show that under fairly general conditions the model with heteroscedastic factors is identified. Identification through heteroscedasticity is also used in King, Sentana, and Waghawani (1994) and Rigobon (2003). The paper by King, Sentana, and Waghawani (1994) is perhaps the closest to ours. Using four observable and two latent heteroscedastic factors the authors
try to explain the links between national stock markets. They show that changes in correlations between markets are driven primarily by movements in latent factors. In contrast, we apply the analysis to the U.S. stock market and focus on sources of systematic risk in industry portfolios.

Apart from Hong, Torous, and Valkanov (2006) our results of predictive regressions are related to the findings of Pollet (2002) and who documents that changes in oil prices can predict the market. We extend his results by showing that returns on other input goods such as gas and metals also have ability to forecast the market and demonstrate that the “input” factor remains significant in predictive regressions even in the presence of returns on input goods.

In the models of Berk, Green and Naik (1999, Gomes, Kogan and Zhang (2003), and Santos and Veronesi (2004), market betas are time-varying. The dynamics of betas in all of the above models are essentially driven by a single factor. In other words, there exists a single source of systematic risk is unique and what changes is firms' exposure to that risk. In our framework there are multiple sources of risk of different nature. While the relative importance of each factor in the market portfolio changes over time, firms' exposure to each type of risk is constant. Since factors correlate with many macroeconomic variables and have different risk-premiums the model implies that cross-sectional dispersion and changes in betas are linked to them as well. Although we do not verify in the current version of the paper, it is quite plausible that the model can deliver similar empirical predictions to those as in Santos and Veronesi (2004).

2.3 Econometric framework

In this section we describe our econometric framework, which involves estimation and identification of common factors.

2.3.1 Model

In our specification, returns obey the following factor model:

\[
\begin{align*}
    r_t &= C f_t + \varepsilon_t, \quad t=1\ldots T \\
    f_t &= \Lambda_t^{1/2} u_t, \quad E(\Lambda_t) = 1 \\
    u_t &\sim N(0, I), \quad \varepsilon_t \sim N(0, \Gamma)
\end{align*}
\] (2.1)

Here \( r_t \) is a \( n \)-vector of returns, \( f_t \) is \( k \)-vector of factors, \( C \) is the \( n \times k \) matrix of loadings, \( \Lambda_t \) is \( k \times k \) diagonal\(^6\) matrix of factor variances at time \( t \), \( I \) is an identity matrix of size \( k \). We assume further that \( E[\varepsilon|f] = 0 \), \( \Gamma \) is \( k \times k \) diagonal matrix of residual variances.

As long as \( \Lambda_t \) varies over time, the model is identified. The identification result holds under very general forms of heteroscedasticity, provided \( \Lambda_{it}, \ i=1\ldots k \) are linearly independent.

\(^6\)Given that we are looking for multiple factors in stock returns that represent exogenous economic shocks, it is natural to assume that the factors are orthogonal.
(Sentana & Fiorentini 2001). Intuitively, requiring factors to be orthogonal at every point in time, results in additional restrictions which serve to identify the model. On the other hand, the special case of constant $\Lambda_t$ over time results in a standard factor model. In this case, it is well-known that matrix of residual variances $\Gamma$ is still identified, but factors are only identified up to an orthogonal rotation matrix $O$. In other words, in the standard factor analysis framework one can identify the linear subspace created by factors but not the factors themselves, which is a substantial drawback of the procedure.

By now, there is a wealth of evidence that stock (Officer 1973, Schwert 1989, ) and bond returns (Schwert 1989, Collin-Dufresne and Goldstein (2002). Similarly, the majority of macroeconomic series are also found to display time-varying volatility (cite verify? Mascaro and Meltzer (1983), Lauterbach (1989)). In fact, Engle’s (1982) seminal ARCH paper uses the heteroscedasticity in the inflation series in the UK as a motivating example. Since stocks are a claim on aggregate output, the real factors that drive output should be reflected in the stock returns as well. This is the main intuition behind the work by Chen, Roll and Ross, and others who link macroeconomic series to stock returns. If these factors represent multiple sources of risk, there is no a-priori reason to believe that volatility across factors is perfectly correlated. Figure 2-1 plots daily stock returns for the NASDAQ and NYSE indices. First, indices display time-varying volatility, and second there appears to be more than one factor that drives volatility.

Factor analysis rests on the assumption of constant loadings, $C$, and our specification preserves this feature. Since we use stock returns to extract factors, some degree of temporal instability perhaps is inevitable. To remedy this problem we work with industry returns rather than with individual stock returns. Industry structure displays more stability than individual firms which may undergo rapid transformations and any change is likely to occur at a low frequency. In addition, Bansal, Fang and Yaron (2004) show that even though the relative shares of industries change over time, this shift has little impact on the moments of stock returns.

### 2.3.2 Estimation

We estimate the model in two steps. In the first stage, we use the limited information likelihood method to estimate the unconditional version of the model. This step does not require any knowledge of $\Lambda_t$ and gives consistent estimates of idiosyncratic variances $\Gamma_k$ and coefficients $C$ up to the rotation matrix $O$. We employ monthly returns to minimize possible market microstructure effects on variance-covariance.

The second stage involves estimation of the rotation matrix $O$. Suppose we start with arbitrary orthogonal factors $\tilde{f}_it$, $i = 1..K$ that belong to the linear space estimated in the first step. Let $\hat{\Lambda}_t$ be their variance-covariance matrix. Identification implies that if factors $\tilde{f}_it$ are the true ones, then $\hat{\Lambda}_t$ is diagonal at all times and vice versa. If $f_t = O\tilde{f}_t$ are new factors...
after an orthogonal transformation, then their variance-covariance matrix is \( \Lambda_t = \mathcal{O}' \hat{\Lambda}_t \mathcal{O} \). The idea then is to find a rotation matrix \( \mathcal{O} \) that produces a diagonal matrix \( \Lambda_t \) for all \( t = 1..T \).

In practice, \( \tilde{\Lambda}_t \) should be estimated. Since the identification result does not depend on the form of heteroscedasticity, we do not need to place unnecessary restrictions on how the factor covariance matrix, \( \Lambda_t \) evolves over time. There are advantages of having the dynamics of \( \Lambda_t \) unspecified. First, as argued by Anderson et al. (2002) and Anderson et al. (2003) in the absence of strong priors on volatility process, realized volatility provides a robust and consistent estimate of factor variability under very general assumptions. Second, by directly estimating volatility, we dramatically simplify estimation of the model. Specifying volatility as a latent process in factor models often leads to econometric specifications that are notoriously difficult to estimate (see, for example, Fiorentini et al 2004).

We estimate \( \tilde{\Lambda}_t \) in monthly increments using a rolling window of daily returns over the past six months:

\[
\hat{\Lambda}_t = \sum_{j=0}^{w} B(j) r_t r_{t-j}.
\]

We use the Bartlett window, \( B(j) \) of width \( w = 2 \) in the estimation of \( \tilde{\Lambda}_t \) out of daily returns to adjust for possible auto(cross)-correlations. As a robustness check, we vary both parameters: window length (three months, one year) and Bartlett window width (3,5,10). In all cases, the results are essentially identical.

Because of the estimation noise, the matrix \( \mathcal{O}' \hat{\Lambda}_t \mathcal{O} \) is unlikely to be diagonal for any choice of the rotation matrix \( \mathcal{O} \). Therefore, to get the estimate of \( \mathcal{O} \) we use a GMM-type estimator. In particular, we choose \( \mathcal{O}^* \) that minimizes the squared off-diagonal elements of \( \mathcal{O}' \hat{\Lambda}_t \mathcal{O} \). That is \( \mathcal{O}^* \) is a solution to

\[
\mathcal{O}^* = \arg \min_{\mathcal{O}} \sum_t \sum_{i \neq j} (\mathcal{O}' \hat{\Lambda}_t \mathcal{O})_{ij}^2.
\]  

As in the case of GMM, one can choose different weights when computing the objective function in (2.2). In our baseline results, we use the correlation matrix of rotated factors

\[
\hat{P}_t \equiv \text{diag}(\mathcal{O}' \hat{\Lambda}_t \mathcal{O})^{-1/2}(\mathcal{O}' \hat{\Lambda}_t \mathcal{O})\text{diag}(\mathcal{O}' \hat{\Lambda}_t \mathcal{O})^{-1/2},
\]

instead of their variance-covariance matrix, \( \mathcal{O}' \hat{\Lambda}_t \mathcal{O} \). Simulations suggest that this results in more robust estimates. Therefore, to conserve the space we report our results only for this case, but both methods produce similar results.
2.3.3 Empirical implementation

We apply our analysis to the set of industry portfolios created by Fama and French. Port-portfolio returns are computed using CRSP data at both daily and monthly frequencies. We focus on the January 1963 through December 2004 period.

Our decision to use industry portfolios is motivated by the assumption that the factor loading matrix, $C$ is time invariant. Industry loadings to macroeconomic factors are likely to be significantly more stable than those of individual stocks. Even though the structure of an industry and thus its response to systematic risk factors may change over time, this change is likely to occur at a much lower frequency than those in individual firms, who may undergo rapid transformations. This also precludes using portfolios of firms sorted on accounting variables such as market equity or book-to-market, given that the economic motivation behind these sorts is not yet well established and as a result there is no a-priori reason to think that they respond in a consistent manner to economic shocks.\footnote{Fama and French classify all stocks into industries based on SIC codes. The portfolio returns as well as classification scheme are available from Kenneth French's Dartmouth web page. In estimation we omit the "Other" industry which is a collection of unsorted firms.}

One has to decide on the number of assets, $N$, to employ in the estimation procedure. We use the 30 industry portfolios, because they are well diversified even in the beginning of our sample, and they are diverse enough to be able to separate different shocks. Nevertheless, our results are similar when we extract factors from the Fama and French 17 or 49 Industry portfolios.

In estimating the $f_t$, we have to decide how much information in the data to use. As in most of the literature, we employ the full sample to estimate the realization of the factors at time $t$, $f_t$. As argued in Ludvigson and Ng (2006), employing the full sample results in the most efficient means of summarizing the covariation in the data, because the estimates do not discard information in the sample. It is even more important in our case since our identification is based on the heteroscedasticity of shocks. If we only used information up to time $t$, the estimates of the factor realizations in the beginning of the sample would be noisier than the estimates towards the end of the sample, thus leading to poor identification. Nevertheless, we show that the analysis produces robust factors. In particular, we demonstrate that factors extracted from the first half of the sample are essentially the same as the ones extracted from the second half.

2.4 Empirical results

2.4.1 The number of factors

The number of common factors in the data is the main parameter of any factor model. The existing literature has proposed a number of tests to determine the number of common factors. Unfortunately, these tests often disagree considerably.\footnote{The industry loadings on our factors are significantly less volatile than those on Fama and French factors, thus alleviating concern raised in the previous literature (Fama and French 1997).}
We use two recent approaches which allow both heteroscedasticity, weak serial and cross-sectional dependence. The first is based on the model selection criteria developed in Bai and Ng (2002). The second uses results from distribution theory of random matrices developed in Onatski (2006a). We refer the reader to those papers for a detailed description of tests. Both tests require an upper bound on the true number of factors \( r_{\text{max}} \). Since there is no theory that would guide us, we use values of \( r_{\text{max}} = 8, 9 \) and \( 10 \) which are suggested in the papers for our dimension of the data. The results are given in Table 2.1. The Table shows that the information criteria of Bai and Ng produce different numbers. On one hand, BIC3, being the most conservative, gives the number of factors equal to one or two. On the other hand, \( PC_i \) criteria, \( i = 1, 2, 3 \), being the most generous, assigns five to eight factors. Onatski's test suggests the number of factors between three and four.

Additionally, our identification technique gives us another way to check the number of factors in the data. To see this, let the true number of factors in the data be \( N \) and suppose that we extract \( M \) factors. It is natural to think that if \( M < N \), the \( M \) extracted factors will be linear combinations of the true \( N \) factors. On the other hand, if \( M > N \), then, given that the model is identified, we expect to see the \( N \) true factors among the extracted factors, with the remaining \( M - N \) being noise. This suggests to start estimation of the model with one factor and then gradually increase the number of factors by one, until the additional factor is not correlated with the previously estimated factors. Although a formal proof for this argument is beyond the scope of the paper, we have performed extensive numerical simulations that provide support\(^{12}\).

In Table 2.2 we present correlations between factors when we estimate the model with three, four, and five factors. We see that when we extract three factors, the first and the third factors spread among the first, third, and fourth factors in a four factor version of the model. At the same time, when we estimate five factors four factors are almost exactly the same as when we estimate the model with four factors, whereas the fifth extracted factor has low correlation with any of the previously estimated factors. This is consistent with the results provided by both tests. If the fifth extracted factor was truly a common factor and not noise, it should be correlated with the previously extracted factors. Therefore, we conclude that there are four common factors in the data.

### 2.4.2 Interpretation

Table 2.3 shows the factor loadings \( C \) and correlation of industry returns with factor returns. We report results when the model is estimated over the whole sample 1963-2004.

Looking at coefficients and correlations, we notice that the first factor correlates highly with all industry returns and can therefore be interpreted as a common market shock. A closer look reveals also that the factor correlates more with industries that are less sensitive to business cycles like Food, Beer, Household, and Health. Looking at how the second factor correlates with industry portfolios, we can see that final consumption good industries like Food, Beer, Smoke, and Utilities have negative correlations. At the same time, industries

\(^{12}\)Available upon request.
that produce capital goods like Fabricated Products, Electrical Equipment, and Business Equipment have positive correlations. This suggests interpretation of the second factor as investment minus consumption industries. Looking at the third factor we can observe that, contrary to the first factor, it correlates more with durable industries whose returns are sensitive to business cycle conditions such as Auto and Clothes, and less with non-durable industries such as Food, Health, and Services. The fourth factor correlates positively with industries whose outputs are raw inputs for production of other goods (Chemicals, Steel, Coal, Oil, Paper) and negatively with other industries. A first interpretation is that the first and third factor captures information about the business cycle, the second factor contains information about the relative price of capital vs consumption goods, whereas the fourth factor contains information about factor input prices.

To verify our initial guess we construct a few auxiliary portfolios. First, we use the 1997 BEA Input-Output tables to construct a portfolio of investment minus consumption industries (IMC). We classify each industry according to their contribution to Consumption Expenditures or Private Investment and then we construct the portfolio as the difference of two value-weighted portfolios for firms that belong to each sector. Second, we use durables minus nondurables good portfolios (DMN) constructed in Gomes, Kogan, and Yogo (2006) which is supposed to capture the risk associated with variations in demand across business cycles. Third, we use the same Input-Output Tables to sort firms into industries that mainly produce raw commodities (Oil, Gas and Metals) versus industries that mainly use these commodities, and then compute the difference of two value-weighted portfolios of each type of firms. We call this portfolio PMU. Lastly, we use commodity prices (oil, gas, and metal prices) to see whether their prices are reflected in our factors.

Table 2.4 presents the correlations of the four extracted factors with the market portfolio, Fama and French's HML and SMB portfolios, the IMC, DMN and PMU portfolios, and returns on input goods. The first factor is highly correlated (85%) with the value-weighted market portfolio, but is essentially uncorrelated with the other portfolios. The third factor is somewhat correlated with the market portfolio (38%), SMB (44%) and DMN (55%). Given that several authors have argued that these portfolios contain information about the business cycle\textsuperscript{13}, this confirms our earlier intuition. The second factor is highly correlated (87%) with investment minus consumption portfolio. Moreover, it is highly negatively correlated (-55%) with the HML portfolio. Papanikolaou (2006) argues that it is not a coincidence and explores this link in more detail. The fourth factor is essentially uncorrelated with all portfolios except ??? and returns on different commodities.

As an additional test, we project our third factor on two durable goods industries (Clothes, Autos) and two non-durable (Services, Health), and our fourth factor on the market portfolio and four input industries (Mines, Steel, Oil and Coal). The results are shown in Table xx. In the first case, the industries enter with the right sign (positive for durables, negative for non-durables), and the correlation of the extracted factor with the fitted values is over 90%. In the second case, the four input industries enter with a positive sign, whereas the market portfolio enters with a negative sign, as expected. The correlation of the fourth

\textsuperscript{13} Vassalou and ? (200x), Gomes, Kogan and Yogo (2006), cite?
factor with its projection is about 85%. These results confirm our initial guess about the economic content of the extracted factors.

2.4.3 Robustness

One of the main concerns of factor analysis is that resulting factors are very sample dependent, and therefore, results estimated in sample may not hold out of sample. To address this issue we divide our sample in two halves and then estimate the model over sub-samples. We then compute the correlation between the extracted factors from each subsample and the factors extracted from the whole sample, shown in Table 2.6. In all cases the correlation is consistently higher than 90%.

One can go a step further and remove sample dependence. As we show above, the second factor is essentially the IMC portfolio, whose construction does not depend on a particular sample. For the other factors we substitute their estimated loadings with their signs. The resulting factors are essentially the same, and the correlations reported in Table 2.7 confirm this. In what follows, we are going to use them as the base factors, since their construction is not sample driven. Table 2.8 shows industry factor loadings and correlations for the newly formed factors. We can see that despite the fact that loadings are now different, the factors have very similar correlation with industry returns to the original ones.

Finally, we can extend our analysis beyond our original sample used to extract the factors. In principle, we can go back as far as 1926, however, at that time the composition of industries are rather slim. Many industries often consist of only two or three firms. As a reasonable compromise we extend our analysis where data permit to the commonly used postwar sample 1951-2004.

2.4.4 Asset pricing tests

Table 2.9 reports summary statistics for the four industry factors. Since the factors are not portfolios but linear combinations of industry excess returns, the means and the standard deviations are not very informative by themselves. We gain a better insight by looking at the factors' Sharpe ratios. The first factor has a large Sharpe ratio, almost twice as high as those of the market portfolio, in the full sample and sub-samples. This is a little bit surprising since it seems to correspond to a less risky part of the market portfolio. Table 2.10 shows that it has positive alpha according to both CAPM and the three Fama and French model. At the same time, the third factor, which supposedly captures the systematic risk associated with cyclical demand for durable goods, has a small Sharpe ratio. From the Table 2.10 we can see that both the CAPM and the three Fama and French model assign it a negative abnormal return. The IMC and the "input" factors have negative Sharpe ratios, which means they have negative risk premiums (estimated using sample average returns). We defer possible explanations for these results to Section (XXX).

14The corresponding results for the original factors are very similar and are reported in Appendix available upon request.
While the risk premium of the first factor is large compared to those of other factors, it would be misleading to conclude from this fact that other factors are unimportant. This is because the reported risk-premiums are unconditional ones. A Table xx shows, returns on our factors are highly predictable ($R^2$ ranges from 1% to 4.7% in monthly regressions), which suggests that the factor risk-premia display a lot of time-variation. Additionally, in almost any equilibrium model, a factor's risk-premium is positively linked to its volatility. When factors are heteroscedastic (see Figure 2-5), average risk-premium might not be large. However, in periods of high factor volatility, the risk-premium can be significantly higher than its average level.

Next, we examine the ability of the four industry factors to explain the cross-section of expected returns. As usual we complement the results with those of the benchmark models: the CAPM and Fama and French three factor model. We begin our analysis of the test of the model by discussing our choice of the cross-section of test assets and a particular form of the test. When factors are excess returns, one can test whether they can price assets using either time-series or cross-sectional tests. In time-series tests, there is an additional restriction of factor risk-premiums being forced to equal the sample average of factor realizations. In cross-sectional tests factor risk-premiums are set in such a way as to minimize cross-sectional pricing errors. This makes cross-sectional tests easier to pass.

First, we sort all NYSE, AMEX and NASDAQ common firms in CRSP into quintiles based on their loadings on our factors. In Table xx we present the time-series intercept of the 5 portfolios sorted on factor $i$ loadings on regressed on returns if the remaining factors $f_{-i}$. In this way, we construct portfolios that have consistently high (and low) loadings on our factors, and then see whether these portfolios earn abnormal returns after controlling for their exposure to the remaining three factors. The pricing errors for all factors, except the first, are monotone in the sort. The fact that sorting on loadings with the first factor does not create a spread in abnormal returns is perhaps not very surprising because (1) there is little dispersion in betas between stocks (verify) (2) it is well known that sorting stocks on their market betas produces no spread in expected returns, and our first factor is highly correlated with the market portfolio.

Next, we use the 48 Fama and French industry portfolios and the 25 Fama and French portfolios sorted on book-to-market and size. It is well-known that both of these cross-sections present, for their respective reasons, a considerable challenge for any of the existing asset pricing models. The 25 Fama and French portfolios are notoriously difficult to price due to the large spread in expected returns (as measured by sample averages) and not so large spread in loadings to most of the factors. To the best of our knowledge, all models to date are rejected when tested on this cross-section, which cause many to judge the success based upon the cross-sectional $R^2$ in cross-sectional tests. This approach was recently critiqued in Daniel and Titman (2005) and Lewellen, Nagel, and Shanken (2005). They show that, because the 25 portfolios have strong factor structure, the $R^2$ is not a very informative statistic and

---

15 The sorting procedure is standard and is described in detail in the Appendix.

16 Fama and French (1993) three factors explain about 90% time-series variation in portfolio returns and about 70% of their cross-sectional variation in their average returns.
the corresponding conclusions should be taken with caution. This problem is not very severe for 48 industry portfolios, which do not have strong factor structure. But, at the same time, they do not have a large spread in their average returns either. So most models usually are not rejected on the basis of cross-sectional tests. The situation is different, however, for the more stringent time-series tests, in which many models show mediocre performance.

As a word of caution, note that in all cases we report the results from unconditional versions of the tests. When factor loadings and factor risk premiums are changing over time, this may lead potentially to a wrong conclusion. However, testing the conditional versions requires modeling of time-variation in factor loadings. However, testing the conditional versions requires modeling of time-variation in factor loadings. A simple possible solution would be to use rolling estimates of factor loading. In our case, however, this does not affect the acceptance or rejection of a model at question. So we do not report the corresponding results.

We use two time-series tests. The first one is Gibbons, Ross, and Shanken (1989) (GRS) test which assumes that pricing errors are i.i.d. over time, homoscedastic and independent of the factors. The second one uses the same test statistics but allows the correction for autocorrelated and heteroscedastic disturbances via GMM. Table 2.11 presents the results of these tests. In all cases the results of both tests agree with each other. Consistent with the previous literature both the CAPM and the three Fama and French factor model are rejected when tested on 48 Fama and French industry portfolios. At the same time, both tests are unable to reject the four factors. Figure 2-2 provides further evidence by showing the pricing errors for the three models across 48 industry portfolios. For almost any portfolio the pricing errors are smaller for the four industry factors than for the CAPM or the three Fama and French factor model. This suggest that the four factors we extracted do indeed capture the risk pertaining to the industry returns. Table 2.10 provides further support for this conclusion. It shows that the extracted industry factors are mispriced by both the CAPM and the three Fama and French model.

When it comes to pricing 25 Fama and French B/M and size portfolios the four industry factors, however, do not show any particular improvement over CAPM in the time-series tests. The reason for this is that, given a strong factor structure of 25 B/M and size portfolios, to successfully price this cross-section, a model has to be able to price HML and SMB Fama and French factors. In fact, since the risk-premium of HML portfolio is almost twice as high as those of SMB it mostly sums up to successful pricing of HML portfolio. While the IMC portfolio has high negative correlation with HML, its sample average return per unit of risk is much lower than those of HML. Figure 2-3 shows the pricing errors for the three models from the GLS and OLS cross-sectional tests. It demonstrates that once risk-premium is allowed to differ from the sample average, the four factors do as well as the three Fama and French model at explaining expected returns on the 25 Fama and French B/M and size portfolios. In particular, the pricing error of the usually mispriced small growth portfolio is much smaller compared to those of CAMP or three Fama and French factor model.

Both Fama and French factors and the four factors we extracted from the industry

---

17See Fama and French (1997) for more results.
portfolios demonstrate that there is a risk factor correlated with B/M. However, the results from tests on the industry portfolios and the 25 B/M and size portfolios suggest that neither HML nor IMC factors are perfect proxies for this risk. While high risk-premium of HML helps to price the 25 B/M portfolios it hurts when it comes to pricing the industry portfolios and the reverse is true for the IMC portfolio.

2.4.5 Predictive regressions

In this section we examine the ability of industry factors to predict future expected returns. Table 2.12 shows cross-(auto)correlations of the factors with three Fama and French factors. We can see that first-order autocorrelations of the third and the fourth factors are 17 % and 16 % respectively which suggests the existence of a persistent component in them. It also makes it plausible for them to predict some other persistent variables, e.g. risk premium. From the table we observe that the third factor has correlation of 15 % with next period Fama and French SMB factor, and the fourth factor has correlation -10 % with next period market portfolio. Since most of the past research was focused on market predictability we concentrate our attention of the market portfolio.

Table 2.13 presents the results of predictive regressions of the excess market return by the fourth factor. The dependent variable is next month’s market return. For each period there are two sets of regression. One is a univariate regression with the fourth factor. In the other we include as a control a number of variables which are known to predict the market: short-term interest rate, yield, spread, and is log of dividend yield. The results, however, do not depend on the inclusion of these variables since they are almost orthogonal to the factor. The regression results show that the fourth factor indeed has the ability to predict the market. A positive shock of one standard deviation to the factor leads to a negative market shock of one standard deviation which is economically significant. We can see that the coefficient at the factor is very stable across different sub-samples. The predictive power is stronger in the second half of the sample, which is again consistent with our interpretation of the fourth factor as a tracking factor for shocks to input prices. Interestingly, as documented in Goyal and Welsh (2005), it is in this sample (the sample does not include 73-74) where most of the known predictors do not perform well.

Our results are related to the findings of Hong, Torous, and Valkanov (2006) who document the ability of industries to lead the market. They find that in the U.S. stock market out of 38 industries 12 industries are significant at 95 % in predictive regressions of the market one month ahead. They also show that similar results hold for the eight largest stock market outside of the U.S. Since in their predictive regressions they include a lagged market portfolio their results essentially show that it is not the industries per se that predict the market but rather some industry components which are orthogonal to the market. Given the ability of the fourth factor to predict the market it is interesting to check, therefore its

---

18 See section Empirical Implementation and Data for more details on their constructions
19 We divide the coefficient from the regression by standard deviations of the fourth factor and the market portfolio.
relation to forecasting ability of the industries. Table 2.14 shows the result of predictive regressions of the next period market excess returns. The regressions are the same as before and similar to those in Hong, Torous, and Valkanov (2006) except that we run a horse race between the fourth factor and 48 Fama and French industry returns. Because of the data availability the sample period is July 1963 to December 2004. As it is evident from the Table 2.14 no industry out of 48 is significant once we include the fourth factor as a control variable. At the same time, the results for the fourth factor are virtually unchanged with regard to inclusion of any specific industry return. The only exception is a lower t-statistic which happens when we include Mines industry.

Another related result is presented in Pollet (2002). He finds that changes in oil prices can also predict stock returns, in particular, the market portfolio. It is interesting to see, therefore, whether the fourth factor has additional predictive power beyond the oil price or, more generally, beyond other commodity prices. Table 2.15 reports the results from predictive regressions identical to those in Table 2.13 but with returns on various commodities. In addition to oil we include return on the composite metal index and gas, which are probably one of the main input goods in the U.S. economy. The sample period of the first two regressions is January 1951 to December 2004. Because meaningful oil and gas monthly prices are available starting 1978 the sample period for the last to regressions is January 1978 to December 2004. The first regression shows that the return on the metal index is a strong predictor of the next period market return. Comparing the result of the second regression in Table 2.15 with those in Table 2.13 we see that inclusion of the metal index almost does not change the coefficient at the fourth factor. Similarly, from the third regression we see that both return on gas and oil have ability to predict the market. The fourth regression confirms that the fourth factor has independent predictive power. Again its coefficient hardly changes.

The reported predictive regressions are in-sample with almost unbelievably high $R^2$ 8.56%. There is an extensive literature which demonstrate that many successful predictors in-sample fail to deliver the same performance out-of-sample\textsuperscript{20}. One of the main indicators for the slashed performance out-of-sample is instability of coefficients when estimated over different sub-samples. As evident, however, from Tables 2.13 and 2.15 the coefficient at the fourth factor is reasonably stable over time which suggests that its ability to predict the market is not sample driven. Table 2.17 provides further evidence on the out-of-sample predictive power of the “input” factor. There we calculate the root mean squared error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (m_t - \hat{m}_t)^2},$$

where $m_t$ is the realized value of the market return and $\hat{m}_t$ is its prediction. We compare two types of forecasts. The first one is based on the historical mean, while the second one is generated by the standard predictive linear regression with $f_{t+1}^k$ on the right hand side. In particular, to get the best estimate of the market return $m_{t+1}$ at time $t$ given the

\textsuperscript{20}See, e. g., Goyal and Welsh (2005) and references therein.
history of the market returns $m_t$ and returns on the "input" factor $f_{4,t}$ up to time $\tau$, we run the regression $m_{t+1} = a + b f_{4,t} + \varepsilon_{t+1}$ and obtain the estimates $\hat{a}$ and $\hat{b}$. Then, the best forecast is $\hat{m}_{\tau+1} = \hat{a} + \hat{b} f_{4,\tau}$. To avoid a look-ahead bias, we estimate the parameters only on the data available to a fictitious observer at the moment $\tau$ and use the estimated parameters to predict the market return one month ahead. Thus, the parameters are re-estimated each month. To form the first forecast, we use the first half of the sample (1951-1977), so the evaluation period starts in 1978. From the Table 2.17 we see that the "input factor" beats the historical average. In particular, the RMSE of the $f_4^*$ is 0.0451 reflecting a noticeable improvement relative to the RMSE of the historical mean which is 0.0453. This error reduction approximately corresponds to the out-of-sample $R^2$ statistic of 0.5% at monthly frequency.

This raises the question of a possible source for the observed predictability. While Pollet (2002) and Hong, Torous, and Valkanov (2006) interpret it in favor of limited information-processing capacity of investors, our results point out toward a more rational explanation. Here we only suggest a tentative story leaving a through modeling for future research. One of the basic ideas in economics and finance is that investment and future stock returns should negatively covary over time. This happens because when the discount rate falls, future expected returns are low, and investment should rise. As we argue the fourth factor traces shocks to input prices. It is plausible, therefore, to assume there is a demand driven component in the "input" factor. Rise in investment, then, leads to higher prices for input goods. So positive returns on the input factor, and more generally, on input goods should predict negative market returns, which is generally consistent with our findings.

2.5 Link to Macroeconomic Variables

In this section we are interested in verifying that our extracted factors indeed capture information about the economy. To this end, we estimate the following model:

$$X_{t+k} - X_t = a_0 + a_1 f_{1,t} + a_2 f_{2,t} + a_3 f_{3,t} + a_4 f_{4,t} + e_t$$

Graph xx plots the coefficients $a(k)$ along with 10% confidence intervals.

In addition, we examine whether our factors predict variables believed to capture the financial investment opportunity set, namely the default premium, dividend yield, yields of 10-y bonds and the short rate. We estimate run the following specification

$$X_t = a_0 + B(L) X_t + a_1 f_{1,t-1} + a_2 f_{2,t-1} + a_3 f_{3,t-1} + a_4 f_{4,t-1} + e_t$$

Here $X_t$ is a vector of financial variables that includes the dividend yield, the short rate, the slope of the yield curve (10-year minus the short rate) and the default spread. The results are shown on Tables xx-xx.

Positive returns on our first factor are followed by an increase in industrial production in

\[\text{21} See, e.g., Lamont (2000).\]
the short to medium run, and a fall in inflation. In the long run however, the first factor seems
to affect only inflation but not industrial production. The fact that this factor predicts prices
and quantities with different signs confirms our earlier intuition that it captures a supply side
shock. This is further supported by the fact that it is correlated with consumer expectations
of future economic conditions in the short run. Furthermore, positive returns on this factor
are followed by a fall in long run yields, probably because of a fall in expected inflation.

Positive returns on our second factor are followed by an increase in inflation. Interestingly,
the predictive power is stronger for PPI, which is consistent with the factor capturing changes
in the relative price of capital goods. In the financial side, it predicts an increase in the slope
of the yield curve, probably because it predicts inflation.

The third factor is a strong predictor of industrial production, both in the short and
the long run, but not inflation. Interestingly, it is strongly correlated with consumer’s
expectations about business conditions, positive returns on the factor are followed by a
fall in default spreads. All of this is consistent with the idea that it captures information
about the business cycle. However, this is information that is somehow orthogonal to the
first factor.

The fourth factor predicts inflation positively, but not industrial production in the short
run, which is consistent with our conjecture that it captures the relative price of input goods.
However, it does predict industrial production in the long run with a negative sign, which
suggests that future output falls because of increased price of input goods. This factor also
predicts higher dividend yield (lower growth rate of dividends), higher slope for the term
structure (again, probably because of inflation), and a fall in the default spread.

2.6 Beta dynamics

It is well-known that market betas of both individual stocks and industry portfolios exhibit
large variation over time\textsuperscript{22}. A multi-factor model with heteroscedastic factors described in
section 2.3.1 suggests a simple explanation for the observed phenomenon. The following
example helps to illustrate the intuition.

Example: Suppose the market portfolio is a sum of two factors

\[ m_t = f_{1,t} + f_{2,t}, \]

and let an asset \( i \) return is given by

\[ R_{i,t} = C_{i,1}f_{1,t} + C_{i,2}f_{2,t} + \epsilon_{i,t}. \]

Then the market beta of the asset \( i \) is

\[ \beta_{i,t} = C_{i,1}\omega_t + C_{i,2}(1 - \omega_t), \quad \omega_t = \frac{\text{Var}_t(f_{1,t})}{\text{Var}_t(f_{1,t}) + \text{Var}_t(f_{2,t})}. \]

\textsuperscript{22}See Fama and French (1997), Andersen et al (2005) among others.
In the example, the market portfolio consists of two factors. The relative amount of risk that a factor contributes to the market portfolio is the ratio of its variance to the total variance of the market portfolio, which we denote as $\omega_t$. If there were only one factor, say $f_{1,t}$, then the asset beta would be just $C_{i,1}$. In a general case it is a convex linear combination of $C_{i,1}$ and $C_{i,2}$. The model also suggests that there should be a co-movement in asset betas of covariances. Of course, since by definition the value-weighted sum of all asset betas equal to one it imposes a restriction on the joint dynamics of asset betas. However, the model predicts more: in the above example the betas for all assets are driven by innovations in only one variable $\omega_t$. More generally speaking, because of the aforementioned restriction on the sum of asset betas, $N$-factor model in returns leads to $(N-1)$-factor model in market betas. In the next sub-section we quantify the ability of our simple model to explain changes in market betas across industry portfolios. Figure 2-4 presents the deviations of market betas measured over different public announcement dates from their unconditional counterparts. Public announcements include GDP (both preliminary, intermediate and advance), CPI (Consumption Price Index), FOMC (Federal Open Market Committee), Non-farm payroll, Construction spending, NAPM (National Association of Purchasing Managers), and are considered to have the largest impact on the stock market. Unconditional betas are estimated by running the regression 2.3 for all trading dates from January 1991 to December 2004 and

$$r_{it} = \alpha_i + \beta_i mkt_t + \varepsilon_i$$

public announcement betas are estimated by running the regression 2.3 on the dates when corresponding public announcements occur. We plot those Fama and French industries (out of 29) whose betas decrease on the FOMC meeting dates. It is quite plausible that information of different nature is released on the dates of public announcements. So the impact of the corresponding news is higher in relative terms on these days as well. Consistent with the model, Figure 2-4 shows a striking comovement of betas across industries. This might have an important implication for estimation of market betas. Comovement of betas means that innovations to betas across different assets are linked to each other. So taking them into account can lead to a more efficient estimate of beta, and therefore, provide a more accurate estimate of the cost of capital.

2.6.1 Empirical Test

In this section we assess the ability of the model to explain time variation in market betas. The analysis of previous sections reveals four factors. In what follows, we take the extracted factors as if they were the true factors, and disregard possible estimation errors of the model at earlier stages. This likely weakens the model’s ability to explain changes in betas over time since estimation errors in factors would lead to time variation in factor loadings, which in turn, would result in poorer fit of the model. Alternatively, we could have a tightly parameterized model from the beginning, in which a time variation in betas can be in statistically rigorous way linked to underlying parameters. We leave it for future research and instead evaluate potential of the model under simplified assumptions. As the
first approximation we assume that the market portfolio can be represented as a linear combination of the four factors

\[ m_t = C_m f_t. \]  

(2.4)

The \( R^2 \) in this regression is 95% which confirms that it is indeed a reasonable approximation\(^{23}\). The betas of industry portfolios are then

\[ \beta = \frac{\text{Cov}(r_t, m_t)}{\text{Var}(m_t)} = \frac{C \Lambda_t C'_m}{C'_m \Lambda_t C_m}. \]

(2.5)

From eq. (2.5) we can see that betas can change either because of change in \( \Lambda_t \), or because of instability of coefficients \( C \) and \( C_m \). In practice there is an additional source of variation – estimation noise since market betas must be estimated. Ideally we would like to separate the contribution of each of these effects to variation in betas. However, this does not seem to be possible on the real data since for that we have to know the true dynamics of betas. As a possible solution we run Monte Carlo simulations on artificial data where we are able to quantify the effect of estimation noise and the true dynamics of betas. The artificial data is constructed to mimic the statistical properties of the real data under the \( H_0 \): that all variation in betas is due to heteroscedasticity of factors. For that the following parameters are estimated from the real data. Factor loadings \( C_m, C \), and variance of residuals \( \Gamma \) are outcome of OLS regressions on factor realizations. For comparability we normalize all industry returns and factors to have a unit variance. It does not change the subsequent results. To capture time variation in volatilities of factors we fit them to GARCH(1,1) model, which is commonly used for such purpose\(^{24}\). The results of estimation are given in Table 2.18 and conditional variances are plotted in Figure 2-5. We can see that the four factors, indeed, exhibit significant heteroscedasticity. The Garch coefficient \( a \) is significantly different from zero for any of the four factors.

The model is then simulated 500 times. To estimate time variation in betas we use three and five-year rolling regressions. The results are almost identical in both cases, so for the sake of space, we present them only for the five-year window estimates. Figure 2-6 reports two 95% simulation bands (solid and dashed lines) of \( \sigma(\beta_t) \) along with the point estimates in the data (stars). The dash line is the 95% simulation band when factors are sampled from a normal distribution with a unit constant variance. Consistent with findings of Fama and French (1997) only five (out of 48) industries fall into the simulated band. The solid line shows the 95% simulation band when factors follow Garch(1,1) process with parameters given Table 2.18. This case represents a significant improvement over the former with now 19 industries falling into the band. It should be noted that industries, which the model falls short to explain, are rather idiosyncratic in nature. For example, they include such industries as Agric, Beer, Toys, Fun, Books, MedEq, Guns, Gold, etc. On the other hand, the model fits well the majority of manufacturing and large industries such as Drugs, Chems, FabPr,

\(^{23}\)A more careful modeling would use the fact that the market portfolio is a value weighted linear combination of industries. If \( S_t \) is a vector of industry market shares then, given the factor model for returns, the market portfolio would be written as \( m_t = S'_t C_m f_t \).

\(^{24}\)See, e.g., Bollerslev, Chou, and Kroner (1992) for a survey.
ElcEq, Autos, Oil, Util, BusSv, and others. In summary, the results support the view that multiple heteroscedastic factors can help to understand the dynamic behavior of the market betas.

2.7 Discussion

Here, we provide a possible explanation for this based upon a version of the $2 \times 2$ production model.

We start with the IMC portfolio. Suppose we have the following production economy with two goods: a consumption good and an investment good. The investment good can be used to produce the consumption good (technology 1) or reinvested to get more of the investment good (technology 2). There are two competitive firms with claims on the output of each technology. Suppose that agents who populate this economy derive their utility only from the consumption good. A positive shock to technology 2, then, leads to an increase in allocation of the investment good to technology 2 and, consequently, cause a decline in today's consumption. Even though the ultimate conclusion depends on particular assumptions about preferences and technologies, it is quite plausible to think that the value of the second firm will rise with respect to the value of the first firm, thus leading to the negative correlation of the IMC portfolio with the consumption growth, and, therefore, to its negative risk-premium. Papanikolaou (2006) formalizes the above intuition in a fully dynamic general equilibrium model.

Let us consider the "input factor". Again, suppose we have the two-good production economy. The second good being now a raw input good which is used to produce the consumption good with a help of some technology, which we assume to be unique for simplicity. There are two competitive firms with claims on the consumption good and the input good respectively. As before, agents who populate this economy derive their utility only from the consumption good. A negative supply shock to the input good, then, leads to an increase in price of the input good and fall in the consumption. As a result the value of the firm with a claim on the input good is negatively correlated with the consumption growth, which justifies the negative risk-premium. It is interesting to notice that the premium of the fourth factor is negative mostly due to the second half of the sample when the volatility of oil shocks greatly increased.

2.8 Conclusions

The volatility of many economic and financial variables changes over time. We take into account the heteroscedasticity in the data and this leads to a complete identification of factors in factor analysis. We demonstrate the value of this methodology by studying the returns of the U.S. industry portfolios. Our analysis reveals four factors with a plausible economic meaning. The factors are shown to be important in explaining the cross-section of expected returns and predicting future returns. The existence of multiple heteroscedastic
factors also provides a simple explanation for the observed volatile behavior of market betas.

We propose a new way to identify economic shocks using the cross-section of stock returns. Our methodology is essentially model-free since it does not rely on a structural model to achieve identification. As a result, it can be applied to other setting as well, for example, to study the comovement in exchange rates across countries, mutual fund styles, etc.
2.9 Tables and Figures

Table 2.1: The number of factors in 29 Fama and French industry portfolios

<table>
<thead>
<tr>
<th>Data Frequency</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Onatski</td>
</tr>
<tr>
<td>Daily</td>
<td>( r_{\text{max}} = 8,9; N=3 )</td>
</tr>
<tr>
<td></td>
<td>( r_{\text{max}} = 10; N=4 )</td>
</tr>
<tr>
<td>Monthly</td>
<td>( r_{\text{max}} = 8; N=3 )</td>
</tr>
<tr>
<td></td>
<td>( r_{\text{max}} = 9; N=4 )</td>
</tr>
</tbody>
</table>

The table reports the number of factors in 29 Fama and French industry portfolios according to the tests developed in Onatski (2006a) and Bai and Ng (2002). There are seven information criteria in Bai and Ng (2002): PCi, ICi, \( i = 1, 2, 3 \), and BIC3. Their results are reported separately only if they give different number of factors. All tests require an upper bound on the true number of factors \( r_{\text{max}} \). Tailored to the dimension of the data, values of \( r_{\text{max}} = 8,9, \text{and} 10 \) are used in the tests. The sample period is January 1963 to December 2004.

Table 2.2: Specification check

<table>
<thead>
<tr>
<th>( \rho(%) )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_3 )</th>
<th>( f_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>90.60</td>
<td>-2.20</td>
<td>27.63</td>
<td>-29.18</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>2.66</td>
<td>99.70</td>
<td>0.87</td>
<td>6.65</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>3.78</td>
<td>-4.83</td>
<td>63.67</td>
<td>75.70</td>
</tr>
<tr>
<td>( f_1 )</td>
<td>99.07</td>
<td>-0.40</td>
<td>-5.07</td>
<td>1.07</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>0.27</td>
<td>98.37</td>
<td>-10.76</td>
<td>1.73</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>2.92</td>
<td>9.57</td>
<td>98.16</td>
<td>-7.65</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>8.80</td>
<td>-8.72</td>
<td>-3.26</td>
<td>82.96</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>8.82</td>
<td>-12.12</td>
<td>-13.06</td>
<td>-55.76</td>
</tr>
</tbody>
</table>

The table reports correlation between extracted factors when the model (2.1) is estimated with three, four, and five factors. The factors are estimated using two step estimation method described in section 2.3.1. The data are returns on 29 Fama and French industry portfolios. The sample period is January 1963 to December 2004.
The table presents loadings of 29 Fama-French industry portfolios on four factors $f_i$, as well as their correlations. The factors are extracted from 29 Fama-French industry portfolios using two step estimation method described in section 2.3.1. Correlations are computed using monthly series over January 1963 to December 2004.
Table 2.4: Correlation of $f_i$ with other returns

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>mkt</th>
<th>smb</th>
<th>hml</th>
<th>IMC</th>
<th>DMN</th>
<th>Oil</th>
<th>Gas</th>
<th>Metal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>0.85</td>
<td>0.04</td>
<td>-0.37</td>
<td>0.22</td>
<td>0.00</td>
<td>-0.09</td>
<td>0.05</td>
<td>-0.03</td>
</tr>
<tr>
<td>$f_2$</td>
<td>0.25</td>
<td>0.40</td>
<td>-0.55</td>
<td>0.87</td>
<td>0.45</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>$f_3$</td>
<td>0.38</td>
<td>0.44</td>
<td>0.18</td>
<td>0.13</td>
<td>0.55</td>
<td>-0.13</td>
<td>-0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>$f_4$</td>
<td>0.05</td>
<td>-0.07</td>
<td>0.19</td>
<td>0.07</td>
<td>-0.07</td>
<td>0.26</td>
<td>0.18</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The table presents correlation of $f_i$ with Fama and French factors, IMC (investment minus consumption) and DMN (durable minus non-durable) portfolios as well as with return on oil, gas and metal indices. The factors $f_i$ are extracted from 29 Fama-French industry portfolios using two step estimation method described in section 2.3.1. Gas, Oil, and Met are returns constructed from Natural Gas Wellhead Price West Texas index, West Texas Intermediate Oil Price, and CRB (BLS) Metals Index correspondingly. IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. DMN is a difference of durables and nondurables portfolios as in Gomes, Kogan, and Yogo (2006). The correlations are computed using monthly series over January 1963 to December 2004, except for oil returns (January 1974 to December 2004) and gas returns (January 1976 to December 2004) because of data availability.

Table 2.5: Factor Interpretation

<table>
<thead>
<tr>
<th>$f_3$</th>
<th>Food</th>
<th>Health</th>
<th>Clothes</th>
<th>Auto</th>
<th>$R^2(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.83</td>
<td>-1.32</td>
<td>4.61</td>
<td>2.85</td>
<td>72.23</td>
</tr>
<tr>
<td></td>
<td>(-5.15)</td>
<td>(-4.68)</td>
<td>(19.25)</td>
<td>(12.93)</td>
<td></td>
</tr>
<tr>
<td>$f_4$</td>
<td>Mkt</td>
<td>Steel</td>
<td>Mines</td>
<td>Coal</td>
<td>Oil</td>
</tr>
<tr>
<td></td>
<td>-7.89</td>
<td>2.68</td>
<td>1.39</td>
<td>1.19</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td>(-22.43)</td>
<td>(12.05)</td>
<td>(9.28)</td>
<td>(10.98)</td>
<td>(15.98)</td>
</tr>
</tbody>
</table>

Table reports coefficients from a projection of factors $f_i$ on a subset of industries, for the period 1963-2004.
Table 2.6: Robustness check

<table>
<thead>
<tr>
<th>ρ(%)</th>
<th>1963 - 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1963-1983</td>
</tr>
<tr>
<td></td>
<td>99.40 1.72  3.92 5.85</td>
</tr>
<tr>
<td></td>
<td>-0.06 -24.04 93.54 10.21</td>
</tr>
<tr>
<td></td>
<td>98.31 -7.82 -4.45 -0.14</td>
</tr>
<tr>
<td></td>
<td>2.32 33.48 90.68 -2.66</td>
</tr>
</tbody>
</table>

The table reports correlation between four extracted factors when the model (2.1) is estimated over different sub-samples. In each case, factors are estimated from 29 Fama and French industry portfolios using two step estimation method described in section 2.3.1.

Table 2.7: Correlation between $f_i$ and $f_i^*$

<table>
<thead>
<tr>
<th>ρ</th>
<th>$f_i^*$</th>
<th>$f_i^*$</th>
<th>$f_i^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>0.95</td>
<td>0.29</td>
<td>-0.01</td>
</tr>
<tr>
<td>$f_3$</td>
<td>0.02</td>
<td>0.86</td>
<td>0.12</td>
</tr>
<tr>
<td>$f_4$</td>
<td>0.17</td>
<td>0.08</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Factors $f_i$, $i = 1, 3, 4$ are estimated from 29 Fama and French industry portfolios using two step estimation method described in section 2.3.1. Factors $f_i^*$, $i = 1, 3, 4$ are constructed by substituting factor loadings in Table 2.3 with their signs. Correlations are computed using monthly series over January 1963 to December 2004.
Table 2.8: Factor loadings and correlation of 29 Fama-French industries for $f_i^*$ and IMC

<table>
<thead>
<tr>
<th>Industry</th>
<th>Loadings</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_i^*$</td>
<td>IMC</td>
</tr>
<tr>
<td>Food</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Beer</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Smoke</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Games</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Books</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Hshld</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Clths</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Hlth</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Chems</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Txtls</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Cnstr</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Steel</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>FabPr</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>ElcEq</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Autos</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Carry</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mines</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Coal</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Oil</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Util</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Telcm</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Servs</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>BusEq</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Paper</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Trans</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Whisl</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rtai</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Meals</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fin</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The table presents loadings of 29 Fama-French industry portfolios on factors $f_i^*$, $i = 1, 3, 4$ and IMC portfolio as well as their correlations. IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. Factor loadings of $f_i^*$, $i = 1, 3, 4$ are obtained by substituting the corresponding factor loadings of $f_i$, $i = 1, 3, 4$ in the Table 2.3 with their signs. Correlations are computed using monthly series over January 1963 to December 2004.
Table 2.9: Summary statistics for $f_i^*$ and IMC

<table>
<thead>
<tr>
<th>Period</th>
<th>$f_i^*$</th>
<th>IMC</th>
<th>$f_3^*$</th>
<th>$f_4^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-1977</td>
<td>8.04</td>
<td>-0.02</td>
<td>0.42</td>
<td>0.16</td>
</tr>
<tr>
<td>1978-2004</td>
<td>8.48</td>
<td>-0.06</td>
<td>1.41</td>
<td>-0.84</td>
</tr>
<tr>
<td>1951-2004</td>
<td>8.26</td>
<td>-0.04</td>
<td>0.92</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

Panel A: Mean (%)

<table>
<thead>
<tr>
<th>Period</th>
<th>$f_i^*$</th>
<th>IMC</th>
<th>$f_3^*$</th>
<th>$f_4^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-1977</td>
<td>45.41</td>
<td>2.62</td>
<td>37.42</td>
<td>24.62</td>
</tr>
<tr>
<td>1978-2004</td>
<td>45.53</td>
<td>4.51</td>
<td>42.35</td>
<td>36.11</td>
</tr>
<tr>
<td>1951-2004</td>
<td>45.44</td>
<td>3.68</td>
<td>39.93</td>
<td>30.88</td>
</tr>
</tbody>
</table>

Panel B: Standard Deviation (%)

<table>
<thead>
<tr>
<th>Period</th>
<th>$f_i^*$</th>
<th>IMC</th>
<th>$f_3^*$</th>
<th>$f_4^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-1977</td>
<td>17.72</td>
<td>-0.66</td>
<td>1.11</td>
<td>0.65</td>
</tr>
<tr>
<td>1978-2004</td>
<td>18.62</td>
<td>-1.33</td>
<td>3.34</td>
<td>-2.33</td>
</tr>
<tr>
<td>1951-2004</td>
<td>18.18</td>
<td>-1.05</td>
<td>2.29</td>
<td>-1.10</td>
</tr>
</tbody>
</table>

Panel C: Sharpe Ratio (%)

Factors $f_i^*$, $i = 1, 3, 4$ are given in Table 2.8. IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. Statistics are calculated from monthly series.
The table reports alpha $\alpha$ and beta of factors $f_i$, $i = 1, 3, 4$ and the IMC portfolio for the CAPM and the three Fama and French model. For comparability a normalized (independent variables are constraint to have the same variance as the market portfolio) $\hat{\alpha}$ is computed. $\hat{\alpha}$ is reported in annualized percentage units. Newey and West (1987) corrected t-statistics are reported in parentheses. Factors $f_i$, $i = 1, 3, 4$ are given in Table 2.8. IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. Statistics are calculated from monthly series.

### Table 2.10: Alphas and betas for $f_i$ and IMC

<table>
<thead>
<tr>
<th>Period</th>
<th>CAPM</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$\beta_{mkt}$</td>
<td>$\hat{\alpha}$</td>
<td>$\alpha$</td>
<td>$\beta_{mkt}$</td>
<td>$\beta_{smb}$</td>
<td>$\beta_{hml}$</td>
<td>$\hat{\alpha}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951-2004</td>
<td>0.03</td>
<td>9.01</td>
<td>3.29</td>
<td>0.03</td>
<td>9.41</td>
<td>-2.42</td>
<td>-0.09</td>
<td>3.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.97)</td>
<td>(21.76)</td>
<td></td>
<td>(3.54)</td>
<td>(28.76)</td>
<td>(-6.47)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.37</td>
<td>-3.59</td>
<td>0.00</td>
<td>0.17</td>
<td>0.49</td>
<td>-0.53</td>
<td>-0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.84)</td>
<td>(8.28)</td>
<td></td>
<td>(-0.21)</td>
<td>(5.45)</td>
<td>(10.49)</td>
<td>(-6.65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>5.91</td>
<td>-3.32</td>
<td>-0.05</td>
<td>6.05</td>
<td>4.78</td>
<td>4.36</td>
<td>-6.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.00)</td>
<td>(13.46)</td>
<td></td>
<td>(-4.91)</td>
<td>(14.30)</td>
<td>(5.12)</td>
<td>(6.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>0.58</td>
<td>-1.13</td>
<td>-0.03</td>
<td>1.49</td>
<td>-0.23</td>
<td>3.95</td>
<td>-4.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.49)</td>
<td>(1.60)</td>
<td></td>
<td>(-2.07)</td>
<td>(4.07)</td>
<td>(-0.36)</td>
<td>(5.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951-1977</td>
<td>0.02</td>
<td>10.26</td>
<td>2.58</td>
<td>0.04</td>
<td>10.40</td>
<td>-1.73</td>
<td>-3.46</td>
<td>4.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td>(17.54)</td>
<td></td>
<td>(3.70)</td>
<td>(22.91)</td>
<td>(-3.01)</td>
<td>(-4.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.21</td>
<td>-2.40</td>
<td>0.00</td>
<td>0.12</td>
<td>0.46</td>
<td>-0.13</td>
<td>-1.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.86)</td>
<td>(6.51)</td>
<td></td>
<td>(-0.68)</td>
<td>(3.40)</td>
<td>(6.10)</td>
<td>(-1.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>6.03</td>
<td>-3.67</td>
<td>-0.05</td>
<td>5.10</td>
<td>6.34</td>
<td>4.84</td>
<td>-6.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.66)</td>
<td>(7.35)</td>
<td></td>
<td>(-4.17)</td>
<td>(9.92)</td>
<td>(9.50)</td>
<td>(6.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.61</td>
<td>-0.33</td>
<td>-0.02</td>
<td>1.25</td>
<td>-2.27</td>
<td>4.55</td>
<td>-3.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.10)</td>
<td>(1.31)</td>
<td></td>
<td>(-1.41)</td>
<td>(2.94)</td>
<td>(-3.06)</td>
<td>(5.56)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978-2004</td>
<td>0.03</td>
<td>8.05</td>
<td>3.98</td>
<td>0.03</td>
<td>8.93</td>
<td>-2.42</td>
<td>1.34</td>
<td>3.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.21)</td>
<td>(17.12)</td>
<td></td>
<td>(2.21)</td>
<td>(22.18)</td>
<td>(-5.14)</td>
<td>(1.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.49</td>
<td>-4.54</td>
<td>0.00</td>
<td>0.15</td>
<td>0.44</td>
<td>-0.77</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.63)</td>
<td>(6.79)</td>
<td></td>
<td>(0.19)</td>
<td>(3.28)</td>
<td>(7.25)</td>
<td>(-7.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>5.82</td>
<td>-2.97</td>
<td>-0.05</td>
<td>6.63</td>
<td>3.76</td>
<td>4.04</td>
<td>-6.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.20)</td>
<td>(12.52)</td>
<td></td>
<td>(-3.06)</td>
<td>(11.10)</td>
<td>(3.07)</td>
<td>(4.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>0.56</td>
<td>-1.81</td>
<td>-0.04</td>
<td>1.97</td>
<td>1.20</td>
<td>4.56</td>
<td>-6.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.55)</td>
<td>(1.06)</td>
<td></td>
<td>(-1.94)</td>
<td>(3.82)</td>
<td>(1.55)</td>
<td>(4.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table reports alpha $\alpha$ and beta of factors $f_i$, $i = 1, 3, 4$ and the IMC portfolio for the CAPM and the three Fama and French model. For comparability a normalized (independent variables are constraint to have the same variance as the market portfolio) $\hat{\alpha}$ is computed. $\hat{\alpha}$ is reported in annualized percentage units. Newey and West (1987) corrected t-statistics are reported in parentheses. Factors $f_i$, $i = 1, 3, 4$ are given in Table 2.8. IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. Statistics are calculated from monthly series.
Table 2.11: Time-series Tests

<table>
<thead>
<tr>
<th>Panel A: 48 Fama and French industry portfolios</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GRS</td>
<td>GMM</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CAPM</td>
<td>1.416</td>
<td>0.040</td>
</tr>
<tr>
<td>3FF</td>
<td>2.504</td>
<td>0.000</td>
</tr>
<tr>
<td>4 factors</td>
<td>0.941</td>
<td>0.589</td>
</tr>
</tbody>
</table>

<p>| Panel B: 25 Fama and French B/M and size |
|------------------------------------------|---|---|</p>
<table>
<thead>
<tr>
<th></th>
<th>GRS</th>
<th>GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPM</td>
<td>4.238</td>
<td>0.000</td>
</tr>
<tr>
<td>3FF</td>
<td>3.088</td>
<td>0.000</td>
</tr>
<tr>
<td>4 factors</td>
<td>4.655</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The table reports the results of two time-series asset pricing tests. The set of test assets includes simple monthly returns of the 48 industry portfolios (Panel A) and the 25 portfolios created by sorting stocks on book to market and market capitalization using NYSE quintiles. The data come from Kenneth French's web-site. The factors considered are the market portfolio, Fama and French's SMB and HML factors, and the four factors constructed from industry portfolios (see Table 2.8). The sample includes monthly data from July 1963 to Dec 2004. The first test is the Gibbons, Ross, and Shanken (1992) (GRS). The second is the GMM time-series test with standard errors corrected for possible autocorrelation and heteroscedasticity. The $S$ matrix is computed using the Newey-West estimator with one lag.

Table 2.12: Factors' Cross-Autocorrelations

<table>
<thead>
<tr>
<th>$\rho(%)$</th>
<th>$f_{1,t}$</th>
<th>IMC$_t$</th>
<th>$f_{3,t}$</th>
<th>$f_{4,t}$</th>
<th>mkt$_t$</th>
<th>smb$_t$</th>
<th>hml$_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{1,t-1}$</td>
<td>-2.18</td>
<td>10.58</td>
<td>13.49</td>
<td>-2.99</td>
<td>4.45</td>
<td>22.65</td>
<td>-3.58</td>
</tr>
<tr>
<td>IMC$_{t-1}$</td>
<td>-4.22</td>
<td>5.08</td>
<td>14.83</td>
<td>-2.03</td>
<td>2.61</td>
<td>6.54</td>
<td>2.45</td>
</tr>
<tr>
<td>$f_{3,t-1}$</td>
<td>-1.18</td>
<td>0.30</td>
<td>17.19</td>
<td>1.79</td>
<td>2.48</td>
<td>15.38</td>
<td>7.11</td>
</tr>
<tr>
<td>$f_{4,t-1}$</td>
<td>-10.79</td>
<td>2.99</td>
<td>-2.62</td>
<td>15.58</td>
<td>-10.00</td>
<td>3.90</td>
<td>5.64</td>
</tr>
</tbody>
</table>

The table reports cross-autocorrelations of the four factors constructed from industry portfolios (see Table 2.8) and three Fama and French factors. Statistics are computed using monthly series over January 1951 to December 2004.
### Table 2.13: Regressions of monthly market excess returns on lagged conditioning variables and $f_A^i$

<table>
<thead>
<tr>
<th>Period</th>
<th>const</th>
<th>$f_A^i$</th>
<th>$r_f$</th>
<th>yield</th>
<th>spread</th>
<th>ldp</th>
<th>$R^2$ (%)</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-2004</td>
<td>0.006</td>
<td>-0.014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.85</td>
<td>647</td>
</tr>
<tr>
<td></td>
<td>(3.370)</td>
<td>(-2.542)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.033</td>
<td>-0.014</td>
<td>-2.447</td>
<td>-0.132</td>
<td>1.833</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.883)</td>
<td>(-2.654)</td>
<td>(-1.704)</td>
<td>(-0.830)</td>
<td>(2.622)</td>
<td>(1.441)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951-1977</td>
<td>0.005</td>
<td>-0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.21</td>
<td>323</td>
</tr>
<tr>
<td></td>
<td>(2.228)</td>
<td>(-1.357)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.033</td>
<td>-0.010</td>
<td>-1.418</td>
<td>-0.597</td>
<td>3.792</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.989)</td>
<td>(-1.328)</td>
<td>(-0.394)</td>
<td>(-1.572)</td>
<td>(2.849)</td>
<td>(0.638)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978-2004</td>
<td>0.006</td>
<td>-0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.07</td>
<td>324</td>
</tr>
<tr>
<td></td>
<td>(2.615)</td>
<td>(-2.152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.175</td>
<td>-0.018</td>
<td>-0.630</td>
<td>-0.890</td>
<td>1.852</td>
<td>0.030</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.198)</td>
<td>(-2.591)</td>
<td>(-0.383)</td>
<td>(-2.999)</td>
<td>(2.093)</td>
<td>(3.694)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table reports estimates from OLS regressions of excess stock returns on lagged variables named in row 1 across different sub-samples. The dependent variable is simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate. The “input” factor $f_A^i$ is described in Table 2.8. $r_f$ is one-month Treasury bill rate. The yield is Moody’s seasoned Aaa corporate bond yield minus one month T-bill rate. The spread is the difference in Moody’s seasoned Aaa and Baa corporate bond yields. ldp is log dividend-price ratio. The dividend-price ratio is computed as dividends over the past year divided by the current price. Newey and West (1987) corrected t-statistics are reported in parentheses. Adjusted $R^2$ and number of observations in each case are displayed.
Table 2.14: Regressions of monthly market excess returns on lagged conditioning variables, $f^*_t$, and industry portfolios

<table>
<thead>
<tr>
<th>Industry</th>
<th>$f^*_t$ coef</th>
<th>t-stat</th>
<th>Ind coef</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agric</td>
<td>-0.017</td>
<td>-2.866</td>
<td>0.033</td>
<td>0.841</td>
</tr>
<tr>
<td>Food</td>
<td>-0.017</td>
<td>-2.643</td>
<td>-0.006</td>
<td>-0.100</td>
</tr>
<tr>
<td>Soda</td>
<td>-0.017</td>
<td>-2.731</td>
<td>-0.006</td>
<td>-0.147</td>
</tr>
<tr>
<td>Beer</td>
<td>-0.017</td>
<td>-2.783</td>
<td>-0.017</td>
<td>-0.393</td>
</tr>
<tr>
<td>Smoke</td>
<td>-0.016</td>
<td>-2.559</td>
<td>-0.032</td>
<td>-1.019</td>
</tr>
<tr>
<td>Toys</td>
<td>-0.017</td>
<td>-2.806</td>
<td>-0.053</td>
<td>-1.601</td>
</tr>
<tr>
<td>Fun</td>
<td>-0.016</td>
<td>-2.731</td>
<td>0.008</td>
<td>0.206</td>
</tr>
<tr>
<td>Books</td>
<td>-0.014</td>
<td>-2.373</td>
<td>0.093</td>
<td>1.531</td>
</tr>
<tr>
<td>Hshld</td>
<td>-0.019</td>
<td>-2.897</td>
<td>-0.075</td>
<td>-1.087</td>
</tr>
<tr>
<td>Clths</td>
<td>-0.016</td>
<td>-2.547</td>
<td>0.026</td>
<td>0.491</td>
</tr>
<tr>
<td>Hlth</td>
<td>-0.017</td>
<td>-2.785</td>
<td>0.000</td>
<td>0.021</td>
</tr>
<tr>
<td>MedEq</td>
<td>-0.018</td>
<td>-2.931</td>
<td>-0.064</td>
<td>-1.067</td>
</tr>
<tr>
<td>Drugs</td>
<td>-0.019</td>
<td>-2.750</td>
<td>-0.063</td>
<td>-0.969</td>
</tr>
<tr>
<td>Chems</td>
<td>-0.016</td>
<td>-2.545</td>
<td>-0.019</td>
<td>-0.267</td>
</tr>
<tr>
<td>Rubbr</td>
<td>-0.017</td>
<td>-2.782</td>
<td>-0.001</td>
<td>-0.013</td>
</tr>
<tr>
<td>Txtls</td>
<td>-0.017</td>
<td>-2.765</td>
<td>0.007</td>
<td>0.167</td>
</tr>
<tr>
<td>BldMt</td>
<td>-0.016</td>
<td>-2.745</td>
<td>-0.014</td>
<td>-0.208</td>
</tr>
<tr>
<td>Cnstr</td>
<td>-0.017</td>
<td>-2.815</td>
<td>0.015</td>
<td>0.368</td>
</tr>
<tr>
<td>Steel</td>
<td>-0.013</td>
<td>-2.007</td>
<td>-0.054</td>
<td>-1.252</td>
</tr>
<tr>
<td>FabPr</td>
<td>-0.016</td>
<td>-2.590</td>
<td>-0.010</td>
<td>-0.277</td>
</tr>
<tr>
<td>Mach</td>
<td>-0.017</td>
<td>-2.539</td>
<td>0.023</td>
<td>0.304</td>
</tr>
<tr>
<td>EleEq</td>
<td>-0.017</td>
<td>-2.776</td>
<td>-0.015</td>
<td>-0.240</td>
</tr>
<tr>
<td>Autos</td>
<td>-0.016</td>
<td>-2.510</td>
<td>-0.046</td>
<td>-0.915</td>
</tr>
<tr>
<td>Aero</td>
<td>-0.017</td>
<td>-2.761</td>
<td>0.032</td>
<td>0.739</td>
</tr>
<tr>
<td>Ships</td>
<td>-0.016</td>
<td>-2.551</td>
<td>-0.018</td>
<td>-0.469</td>
</tr>
<tr>
<td>Guns</td>
<td>-0.017</td>
<td>-2.630</td>
<td>0.001</td>
<td>0.039</td>
</tr>
<tr>
<td>Gold</td>
<td>-0.015</td>
<td>-2.215</td>
<td>-0.016</td>
<td>-0.625</td>
</tr>
<tr>
<td>Mines</td>
<td>-0.011</td>
<td>-1.563</td>
<td>-0.072</td>
<td>-1.719</td>
</tr>
<tr>
<td>Coal</td>
<td>-0.016</td>
<td>-2.348</td>
<td>-0.004</td>
<td>-0.127</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.014</td>
<td>-2.107</td>
<td>-0.046</td>
<td>-0.815</td>
</tr>
<tr>
<td>Util</td>
<td>-0.019</td>
<td>-3.332</td>
<td>0.121</td>
<td>1.939</td>
</tr>
<tr>
<td>Telem</td>
<td>-0.018</td>
<td>-2.999</td>
<td>-0.054</td>
<td>-0.826</td>
</tr>
<tr>
<td>PerSv</td>
<td>-0.016</td>
<td>-2.748</td>
<td>0.060</td>
<td>1.460</td>
</tr>
<tr>
<td>BusSv</td>
<td>-0.016</td>
<td>-2.516</td>
<td>0.035</td>
<td>0.426</td>
</tr>
<tr>
<td>Comps</td>
<td>-0.018</td>
<td>-2.958</td>
<td>-0.032</td>
<td>-0.685</td>
</tr>
<tr>
<td>Chips</td>
<td>-0.016</td>
<td>-2.737</td>
<td>0.026</td>
<td>0.517</td>
</tr>
<tr>
<td>LabEq</td>
<td>-0.017</td>
<td>-2.794</td>
<td>0.013</td>
<td>0.238</td>
</tr>
<tr>
<td>Paper</td>
<td>-0.017</td>
<td>-2.761</td>
<td>0.003</td>
<td>0.047</td>
</tr>
<tr>
<td>Boxes</td>
<td>-0.017</td>
<td>-2.772</td>
<td>-0.006</td>
<td>-0.116</td>
</tr>
</tbody>
</table>
Table 2.14: (Continued): Regressions of monthly market excess returns on lagged conditioning variables, $f_d$, and industry portfolios

<table>
<thead>
<tr>
<th>Industry</th>
<th>$f_d$</th>
<th>Ind</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef</td>
<td>t-stat</td>
</tr>
<tr>
<td>Trans</td>
<td>-0.017</td>
<td>-2.791</td>
</tr>
<tr>
<td>Whlsl</td>
<td>-0.017</td>
<td>-2.812</td>
</tr>
<tr>
<td>Rtail</td>
<td>-0.015</td>
<td>-2.272</td>
</tr>
<tr>
<td>Meals</td>
<td>-0.017</td>
<td>-2.689</td>
</tr>
<tr>
<td>Banks</td>
<td>-0.016</td>
<td>-2.686</td>
</tr>
<tr>
<td>Insur</td>
<td>-0.016</td>
<td>-2.703</td>
</tr>
<tr>
<td>RlEst</td>
<td>-0.017</td>
<td>-2.775</td>
</tr>
<tr>
<td>Fin</td>
<td>-0.016</td>
<td>-2.798</td>
</tr>
<tr>
<td>Other</td>
<td>-0.017</td>
<td>-2.804</td>
</tr>
</tbody>
</table>

The table reports estimates from OLS regressions of excess stock returns on lagged variables named in column 1. The dependent variable is simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate. The "input" factor $f_d$ is described in Table 2.8. $r_f$ is one-month Treasury bill rate. The yield is Moody's seasoned Aaa corporate bond yield minus one month T-bill rate. The spread is the difference in Moody's seasoned Aaa and Baa corporate bond yields. Idp is log dividend-price ratio. The dividend-price ratio is computed as dividends over the past year divided by the current price. Newey and West (1987) corrected t-statistics are reported in parentheses. Adjusted $R^2$ and number of observations in each case are displayed. The sample period is July 1963 to December 2004.
Table 2.15: Regressions of monthly market excess returns on lagged conditioning variables, $f_4^*$, and commodity returns

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.033</td>
<td>0.033</td>
<td>0.166</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(1.85)</td>
<td>(3.73)</td>
<td>(3.89)</td>
</tr>
<tr>
<td>$r_f$</td>
<td>-2.795</td>
<td>-2.800</td>
<td>-0.770</td>
<td>-0.677</td>
</tr>
<tr>
<td></td>
<td>(-2.00)</td>
<td>(-2.01)</td>
<td>(-0.46)</td>
<td>(-0.41)</td>
</tr>
<tr>
<td>yield</td>
<td>-0.071</td>
<td>-0.082</td>
<td>-0.808</td>
<td>-0.873</td>
</tr>
<tr>
<td></td>
<td>(-0.46)</td>
<td>(-0.54)</td>
<td>(-2.72)</td>
<td>(-2.93)</td>
</tr>
<tr>
<td>spread</td>
<td>0.016</td>
<td>0.016</td>
<td>0.017</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(2.29)</td>
<td>(2.35)</td>
<td>(1.87)</td>
<td>(1.97)</td>
</tr>
<tr>
<td>ldp</td>
<td>0.007</td>
<td>0.007</td>
<td>0.029</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(1.43)</td>
<td>(1.41)</td>
<td>(3.30)</td>
<td>(3.40)</td>
</tr>
<tr>
<td>$f_4^*$</td>
<td>-0.011</td>
<td>-0.016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gas</td>
<td></td>
<td>0.041</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.23)</td>
<td>(2.55)</td>
<td></td>
</tr>
<tr>
<td>oil</td>
<td></td>
<td>-0.099</td>
<td>-0.088</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-3.2)</td>
<td>(-2.85)</td>
<td></td>
</tr>
<tr>
<td>met</td>
<td>-0.138</td>
<td>-0.119</td>
<td>-0.084</td>
<td>-0.052</td>
</tr>
<tr>
<td></td>
<td>(-2.99)</td>
<td>(-2.50)</td>
<td>(-1.28)</td>
<td>(-0.75)</td>
</tr>
<tr>
<td>$R^2(%)$</td>
<td>3.700</td>
<td>4.19</td>
<td>7.49</td>
<td>8.59</td>
</tr>
<tr>
<td>$T$</td>
<td>647</td>
<td>647</td>
<td>324</td>
<td>324</td>
</tr>
</tbody>
</table>

The table reports estimates from OLS regressions of excess stock returns on various industry portfolio returns, the “input” factor $f_4^*$, and other lagged variables. The dependent variable $m_{t+1}$ is simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate. Only the coefficients in front of the lagged industry return and the “input” factor $f_4^*$ are reported. The “input” factor $f_4^*$ is described in Table 2.8. Other forecasting variables include one-month Treasury bill rate $r_f$, the yield (Moody’s seasoned Aaa corporate bond yield minus one month T-bill rate), the spread (the difference in Moody’s seasoned Aaa and Baa corporate bond yields), ldp (log dividend-price ratio, the dividend-price ratio is computed as dividends over the past year divided by the current price), gas, oil, and met are returns constructed from Natural Gas Wellhead Price West Texas index, West Texas Intermediate Oil Price, and CRB (BLS) Metals Index correspondingly. The indices are available from the Global Financial database. Newey and West (1987) corrected t-statistics are reported in parentheses. The sample period is January 1951 to December 2004 for the first two regressions and January 1978 to December 2004 for the last two regressions.
Table 2.16: Regressions of market excess returns on lagged $f_1$

<table>
<thead>
<tr>
<th>Coef.</th>
<th>$t$-stat</th>
<th>$R^2%$</th>
<th>Coef.</th>
<th>$t$-stat</th>
<th>$R^2%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.013</td>
<td>-2.37</td>
<td>0.85</td>
<td>-0.014</td>
<td>-2.00</td>
</tr>
<tr>
<td>2</td>
<td>-0.022</td>
<td>-2.41</td>
<td>1.08</td>
<td>-0.020</td>
<td>-1.77</td>
</tr>
<tr>
<td>3</td>
<td>-0.023</td>
<td>-2.13</td>
<td>0.76</td>
<td>-0.028</td>
<td>-2.06</td>
</tr>
<tr>
<td>12</td>
<td>-0.062</td>
<td>-2.10</td>
<td>1.36</td>
<td>-0.069</td>
<td>-3.05</td>
</tr>
<tr>
<td>24</td>
<td>-0.085</td>
<td>-2.82</td>
<td>1.39</td>
<td>-0.104</td>
<td>-3.46</td>
</tr>
<tr>
<td>36</td>
<td>-0.033</td>
<td>-1.19</td>
<td>0.01</td>
<td>-0.070</td>
<td>-2.55</td>
</tr>
</tbody>
</table>

The table reports estimates from OLS regressions of excess stock returns on the lagged “input” factor $f_1$, and other lagged variables. The dependent variable $m_{t+1}$ is simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate. The “input” factor $f_1$ is described in Table 2.8. Newey and West (1987) corrected $t$-statistics are reported in parentheses. The sample period is January 1951 to December 2004 for the first regression and January 1978 to December 2004 for the second regression.

Table 2.17: Out-of-sample forecasting power of $f_4$

<table>
<thead>
<tr>
<th>Predictive variable</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical average of returns</td>
<td>0.0453</td>
</tr>
<tr>
<td>$f_4$</td>
<td>0.0451</td>
</tr>
</tbody>
</table>

Table reports a root mean squared error of prediction (RMSE) of the historical mean and the “input” factor $f_4$ described in Table 2.8. $RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (m_t - \hat{m}_t)^2}$, where $m_t$ is a simple return on the CRSP value-weighted stock market index over one-month Treasury-bill rate and $\hat{m}_t$ is its prediction. A forecast $\hat{m}_{t+1}$ is given by $\hat{m}_{t+1} = \hat{a} + \hat{b} f_{4,t}$, where the estimates $\hat{a}$ and $\hat{b}$ are obtained from the regression $m_{t+1} = a + b f_{4,t} + \epsilon_{t+1}$, $t + 1 \leq \tau$. To form the first forecast, the first half of the sample (1951-1977) is used, so the evaluation period starts in January 1978 and ends in December 2004.
Table 2.18: GARCH(1,1) estimates of $f_i$

<table>
<thead>
<tr>
<th>Garch(1,1)</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$f_3$</th>
<th>$f_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.091</td>
<td>0.023</td>
<td>0.049</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.012)</td>
<td>(0.017)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$a$</td>
<td>0.099</td>
<td>0.131</td>
<td>0.129</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>$b$</td>
<td>0.810</td>
<td>0.838</td>
<td>0.825</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.041)</td>
<td>(0.035)</td>
<td>(0.030)</td>
</tr>
</tbody>
</table>

Table reports estimates of GARCH(1,1) parameters for the factors are extracted from 29 Fama-French industry portfolios using two step estimation method described in section 2.3.1 (see also Table 2.3). Each factor is demeaned and is fitted to the following model: $f_{i,t} = \sigma_{i,t} \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is an independent, identically distributed (i.i.d.) sequence $\sim N(0,1)$. The time-conditional variance, $\sigma_{i,t}^2$, process is modeled as $\sigma_{i,t}^2 = \kappa + b \sigma_{i,t-1}^2 + a \varepsilon_{i,t}^2$. Standard errors are provided in parenthesis. Estimation is carried by using garchfit routine included in the software package MATLAB.
The figure plots daily stock returns for the NASDAQ and NYSE indices. The sample period is January 1973 to December 2004.
Figure 2-2: Pricing errors. 48 Fama and French industry portfolios.

The figure pictures the pricing errors $\alpha_i$ of the 48 Fama and French industry portfolios from OLS time-series regressions: $r_{it} = \alpha_i + \beta_i^F F_t + \epsilon_i^t$. $r_{it}$ is the industry excess return over one-month Treasury bill rate. The factors $F_t$ are simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate (CAPM), three Fama and French factors (3FF), and the four factors $f_1^i$, IMC, $f_3^i$ and $f_4^i$ (4 factors). $f_1^i$, $f_3^i$ and $f_4^i$ are excess returns constructed out of 29 Fama and French industry portfolios (see Table 2.8). IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. $\alpha_i$ are annualized. The sample period is July 1963 to December 2004.
Figure 2-3: Pricing errors. 25 Fama and French B/M and Size sorted portfolios.

The figure pictures the pricing errors $\alpha_i$ of the 25 Fama and French B/M and size sorted portfolios from GLS (Upper panel) and OLS (Bottom panel) cross-sectional regressions: $E(r_{ut}) = \alpha_i + \beta_i \lambda$. $E(r_{it})$ is the expected industry excess return over one-month Treasury bill rate. The factors are simple return on the CRSP value-weighted stock market index over the one-month Treasury-bill rate (CAPM), three Fama and French factors (3FF), and the four factors $f_1^i$, IMC, $f_3^i$ and $f_4^i$ (4 factors). $f_1^i$, $f_3^i$ and $f_4^i$ are excess returns constructed out of 29 Fama and French industry portfolios (see Table 2.8). IMC is a portfolio of stocks producing investment goods minus stocks producing consumption goods. $\alpha_i$ are annualized. The sample period is July 1963 to December 2004.
The figure pictures deviations of market betas measured over different public announcement dates from their unconditional counterparts for a subset of 29 Fama and French industry portfolios. Public announcement dates are: 1 - GDP (Gross Domestic Product), 2 - CPI (Consumption Price Index), 3 - FOMC (Federal Open Market Committee), 4 - Non-farm payroll, 5 - Construction spending, 6 - NAPM (National Association of Purchasing Managers). Betas are estimated from the regression: $r_{it} = \alpha_i + \beta_i mkt_t + \epsilon_{it}^t$ (*) using daily returns. Unconditional betas are estimated by running the regression (*) for all trading dates from January 1991 to December 2004. Public announcement betas are estimated by running the regression (*) on the dates when corresponding public announcement occurs. The industries are chosen are those whose betas decline on FOMC public announcement dates.
Figure 2-5: Conditional Variance of \( f_i \)

The figure plots conditional variances of the four factors \( f_i, \ i = 1..4 \), extracted from 29 Fama-French industry portfolios using two step estimation method described in section 2.3.1 (see also Table 2.3). Each factor is demeaned and is fitted to the following model:

\[
 f_{i,t} = a_i t e_{i,t},
\]

where \( e_{i,t} \) is an independent, identically distributed (i.i.d.) sequence \( \sim N(0,1) \). The time-conditional variance, \( \sigma^2_{i,t} \), process is modeled as

\[
 \sigma^2_{i,t} = \kappa + b \sigma^2_{i,t-1} + a e^2_{i,t}.
\]

Corresponding estimates of \( \kappa, b, \) and \( a \) are given in Table 2.18.
The figure plots standard deviation of market betas for the 48 Fama and French industry portfolios (stars) and 95 % simulation bands of market beta standard deviations from the model described in Section 2.6.1. The dash line is 95 % simulation band when factors are sampled from a normal distribution with a unit constant variance. The solid line shows a 95 % simulation band when factors follow Garch(1,1) process with pre-estimated parameters given Table 2.18.
Chapter 3

Financial Relationships and the Limits to Arbitrage

This chapter is based on work with Jiro E. Kondo.¹

Abstract

In this chapter we propose a new foundation for the limits to arbitrage based on financial relationships between arbitrageurs and banks. Financially constrained arbitrageurs may choose to seek additional financing from banks who can understand their strategies. However, a hold-up problem arises because banks cannot commit to provide capital and have the financial technology to profit from the strategies themselves. Wary of this, arbitrageurs will choose to stay constrained and limit their correction of mispricing unless banks have sufficient reputational capital. Using the framework of stochastic repeated games, we show that this form of limited arbitrage arises when mispricing is largest and becomes more substantial as the degree of competition between banks intensifies and arbitrageur wealth increases.

Having worked at a major Wall Street bank, [John Meriwether] felt that investment banks were rife with leaks and couldn’t be trusted not to swipe his trades for themselves.

- Lowenstein (Ch. 3, 2000)

Because they act as a gateway to the entire range of services investment banks offer, prime brokers gain access to privileged market-sensitive information from their clients - information that must be kept from other clients and from the investment bank’s proprietary trading departments.

- Financial Times²

¹The authors would like to thank Patrick Bolton, Kevin Chu, Alex Edmans, Ilan Guedj, Steve Ross, David Scharfstein, Eric Van den Steen and participants at the MIT Finance Lunch for helpful comments. Special thanks to Dimitri Vayanos for his encouragement and guidance during the early stages of this project.

3.1 Introduction

Arbitrageurs rely on investment banks for a substantial portion of their financing. However, when accessing this source of finance, they also run the risk of being expropriated through activities such as front-running. So why do they use this source of funding instead of other forms of external finance? First, we argue that investment banks face less severe adverse selection problems. Namely, they are a type of informable financier because arbitrageurs can reveal their private information to them. Second, a bank’s reputation can also allow it to credibly commit to not exploit arbitrageurs because she values her financial relationships with them. However, this commitment is limited. This chapter studies the role of financial relationships in enabling informable finance and proposes a new foundation for the limits to arbitrage.

Using a stochastic repeated game that arises from time-variation in mispricing, we develop further predictions about when arbitrage becomes limited. Namely, we show that limited arbitrage occurs at times when mispricing is most severe. This observation cannot be obtained from standard repeated game models and is robust to realistic assumptions about contractibility. We also demonstrate that, under certain conditions, our limits to arbitrage problem becomes more substantial as competition increases between investment banks in attracting business from arbitrageurs and, surprisingly, as arbitrageur wealth increases. This last finding can partially explain why arbitrageurs occasionally choose to refuse additional capital from desirable institutional investors.

Existing models of limited arbitrage typically introduce a source of non-fundamental risk and conclude that arbitrage opportunities are risky investments (e.g., De Long et al. (1990) and Shleifer-Vishny (1997)). In these settings, arbitrageurs are reluctant to trade aggressively against mispricing out of fear that it will worsen and lead them to liquidate their positions at a loss. This is motivated by two assumptions: arbitrageurs are subject to financial constraints and potential capital providers cannot understand the arbitrageur’s investment opportunities. The latter assumption is called the “separation of brains and capital.”

Yet, the view that all potential financiers cannot understand the arbitrageurs’ strategies is extreme. Though arbitrageurs have significantly more expertise than the typical investor, some knowledgeable providers of capital have the experience and skill required to adequately evaluate their opportunities. For instance, investment banks gain similar expertise through proprietary trading while also funding arbitrageurs through prime brokerage operations. In the presence of such informable finance, one might think that arbitrage would not be limited. Regardless of recent performance, arbitrageurs could exploit new opportunities by simply revealing their strategies to banks and then borrowing funds or getting them to reduce margin requirements.

However, this logic is flawed. Since courts, like most investors, cannot understand the content of communication between the two parties, contracts are incomplete and banks cannot commit through formal contracts to provide any capital ex-post. Moreover, the experience that allows banks to understand the arbitrageurs also gives them access to similar, if not superior, financial technology to execute these strategies. This creates a hold-up
problem. After the arbitrageur reveals his information to the bank, what prevents the latter from providing no capital and undertaking the profitable transactions for herself?³

In a one-shot transaction, the answer is nothing. Banks do not lose anything when holding up arbitrageurs and therefore cannot credibly commit to making arbitrageurs better off when they reveal their strategies. Arbitrageurs understand this and refuse to share information with them. Given this decision, their only resort is to seek financing from the uninformed (à la Shleifer-Vishny). The separation of brains and capital arises endogenously.

However, when similar interactions occur repeatedly, a bank’s concern for its reputation can persuade the arbitrageur to reveal his information. We consider a sequence of arbitrageurs, each with knowledge of a different arbitrage opportunity, who interact with infinitely lived banks. In equilibrium, arbitrageurs and banks optimally collude whenever the latter can commit to acceptable behavior. Specifically, arbitrageurs reveal their information and allow banks to keep some profits for themselves in such a way that total surplus is maximized.

Nevertheless, the power of financial relationships as a disciplinary mechanism is limited. The value of relationships is proportional to the average profitability of an arbitrage opportunity whereas a bank’s profits from fully expropriating the arbitrageur varies with the current level of mispricing. If the most profitable arbitrages are sufficiently superior to the average one, banks won’t always be able to commit to cooperation. This problem occurs exactly when mispricing is largest and communication would create the largest surplus between the two parties.

In addition to the limits to arbitrage literature, our paper is also related to several other strands of finance. First, it provides an additional example where some investors can take advantage of others through knowledge of their proprietary strategies and trading needs. This is related to the work of Brunnermeier-Pedersen (2004) who study predatory trading in response to the predictable activity of large investors. Ko (2002) considers front-running by banks as a cost to arbitrageurs in disclosing their risk profiles and shows that this can lead to endogenous concentrations of risk. His analysis differs from ours because it is static and relies on assumptions that arbitrageurs’ strategies are risky and banks are risk-averse. Finally, our work is also related to the theoretical literature on the efficiency benefits of institutional reputation in finance (e.g., Sharpe (1990), and Chemmanur-Fulghieri (1994a, 1994b)), ⁴

This chapter is also closely related to the general setting of financing innovation and selling ideas when intellectual property rights are imperfect. Because of informational asymmetries, potential financiers and buyers are unlikely to offer a fair price for valuable innovations and good ideas unless details are provided to them ex-ante. However, once they have this knowledge, they may effectively own all its productive use and have little incentive to pay for it ex-post. This hold-up problem, which is identical to ours, is known as the fundamental

³More precisely, as alluded to earlier in the Financial Times quote, information provided by the arbitrageur to the bank’s prime brokerage division can be leaked to the firm’s proprietary traders.

⁴Our model is also similar to Rotemberg-Saloner (1986) who study a stochastically repeated game with i.i.d. variation. However, all the strategic players in their setting are homogeneous while we assume heterogeneity between the arbitrageur and the bank in the form of initial information asymmetry.
paradox (Arrow (Ch. 6, 1971)).

There is a large literature that explores ways of mitigating this problem. Anton-Yao (1994) show that the existence of competition among potential buyers can improve efficiency because the entrepreneur can threaten to reveal his idea to a competitor. Rajan-Zingales (2001) study how organizational hierarchy can be used to minimize the problem of information leakage. Rather than looking at commitment, Anton-Yao (2002, 2004) explore the use of partially expropriable disclosures to signal project value. Nevertheless, none of these models achieve the first-best. In independent work, Hellman-Perotti (2005) show that firm reputation can foster more efficient innovation. A reinterpretation of our model adds to their observation by remarking that the reputation mechanism is limited in a particular way: it fails to achieve first-best implementation in the most valuable cases.

This chapter has two main contributions. First, it provides new economic foundations for the limits to arbitrage: fear of opportunism by informable financiers. This complements existing theory by identifying conditions under which the separation of brains and capital would endogenously arise. The second contribution is broader. Namely, this chapter is one of the first to explicitly incorporate the notion of financial relationships in the asset pricing and microstructure literature. Other work includes Benabou-Laroque (1992), Desgranges-Foucault (2002), and Bernhardt et al. (2004). The observation that trade and financial interaction among agents, especially large institutions, is not anonymous is obvious, yet its effect on pricing has remained largely unexplored. It would be interesting to investigate how this aspect of trade ameliorates or worsens informational asymmetries and incentive problems in financial markets. Such analysis could also lead to a better understanding of the existing institutional structure of the securities industry. We open a discussion on the market structure of arbitrage later in this chapter by commenting on the role that certain arbitrage institutions, like fund-of-funds and seeders, play in overcoming our hold-up problem.

The remainder of this chapter proceeds as follows. Section 2 presents the model and its assumptions. Section 3 derives a characterization of the equilibria and their general properties. Section 4 considers a simple refinement based on the degree of banking competition and presents examples of its implication on limited arbitrage. Additional extensions and discussion are in section 5. All proofs are gathered in the Appendix.

3.2 The Model

We consider a parsimonious framework where arbitrage opportunities are riskless and converge immediately following a round of trading. An arbitrage opportunity is modeled as an extensive form game where initial mispricing is generated at $t = 0$, relational interactions between the arbitrageur and the investment bank (bank) occur at $t = 1$, one round of trading takes place at $t = 2$, and terminal payoffs are realized at $t = 3$. In this section, all variables, including trading profits, are observable but not verifiable. We relax this assumption in section 5.1 and show that our results remain largely unaffected by contracting if we allow for adverse selection. We study both the one-shot and infinitely repeated versions of this game. The stage game in the latter version is called an arbitrage opportunity cycle.
There are $N$ risky assets in the economy and a riskless asset with a rate of return normalized to 0. All agents have a common prior over the terminal payoffs of these assets, namely that they are imperfectly correlated with identical means, $\bar{\nu}$. The arbitrageur receives a private signal at $t = 0$, informing him that two assets, $A_1$ and $A_2$, have identical terminal payoffs. We will focus on equilibrium in the markets for these two assets. For simplicity, we assume that the universe of assets is sufficiently large that all other investors cannot infer the arbitrageur’s information from prices as the Bayesian probability that a given asset is part of the arbitrageur’s signal is always negligible.

In each market, there are two types of non-strategic investors who place demands: noise traders and long-term traders. Noise traders buy and sell randomly in each market and are responsible for the existence of arbitrages. Their trading in $A_1$ and $A_2$ is $X_{N1}$ and $X_{N2}$, respectively. We denote by $F$ the cumulative distribution of the noise trader spread, $\Delta X_N = |X_{N1} - X_{N2}|$, and assume that it has a finite second moment. In this environment, $\Delta X_N$ can be seen as a measure of initial mispricing. Long-term traders submit downward sloping linear demands for each risky asset,

$$X_{L_{Ri}} = \frac{1}{\lambda}(\bar{\nu} - p_t),$$

where $\lambda > 0$. This specification of residual demand curves is standard in the literature (e.g., Brunnermeier-Pedersen (2004) and Xiong (2001)). Various interpretations can be given to the long-term traders. For example, they can be viewed as market makers à la Grossman-Miller (1988) or uninformed investors fearing exploitation by informed ones (e.g., Grossman-Stiglitz (1980) and Kyle (1985)).

The main analysis focuses on the behavior of a sequence of short-lived arbitrageurs and a long-lived bank. Both groups are risk-neutral and strategic. The bank lives forever and has a discount factor $\delta$.\(^5\) Arbitrageurs live for one cycle and place convergence trades on $A_1$ and $A_2$.\(^6\) They have limited wealth and face a default financial constraint of the form,

$$X_A \leq M_L,$$

where $X_A$ is the amount of convergence trading undertaken by the arbitrageur. This constraint captures the arbitrageurs’ limited access to initial funding from both investors and banks, the latter through margin financing. The microeconomic foundations of this type of constraint are well understood (e.g., Kiyotaki-Moore (1997) and Gromb-Vayanos (2002)).

In each arbitrage opportunity cycle, the arbitrageur observes $\Delta X_N$ and can trade without revealing his strategy to the bank. This choice is denoted by $R = 0$. The bank does not

\(^6\)The discount factor needn't only be a proxy for bank impatience. It can also be seen as a proxy for the frequency of discovering new arbitrages or as a reduced form for relevant elements that aren't modeled here, like bank risk aversion or agency problems within the bank.

\(^6\)A convergence trade is defined as a long position in $A_i$ and an equally short position in $A_j$. In a different setting, Xiong (2001) also studies equilibrium mispricing under the same assumption that arbitrageurs are restricted to such trades.
trade in this case and prices are determined by market clearing:

\[ p_1 = \bar{v} + \lambda (X_{N1} - X_A) \quad \text{and} \quad p_2 = \bar{v} + \lambda (X_{N2} + X_A). \]  

(3.3)

Alternatively, the arbitrageur can choose to share his information to the bank in the hope of negotiating an increase in his trading capacity. This increase can be achieved with an infusion of capital or, as is more common, a renegotiation of terms in the margin agreement. This choice is denoted by \( R = 1 \). Unfortunately, the bank cannot commit ex-ante to alleviate the arbitrageur’s financial constraint and may select any position limit satisfying \( M \geq M_L \). It can also choose to front-run the arbitrageur by trading an amount \( X_B \geq 0 \) ahead of him. The arbitrageur observes \( X_B \) prior to choosing his own position, \( X_A \leq M \) and all trades clear simultaneously at the market clearing price:

\[ p_1 = \bar{v} + \lambda (X_{N1} - X_A - X_B) \quad \text{and} \quad p_2 = \bar{v} + \lambda (X_{N2} + X_A + X_B). \]  

(3.4)

Profits for the arbitrageur and bank are given by \( \Delta p \cdot X_A \) and \( \Delta p \cdot X_B \), respectively, where \( \Delta p \equiv p_1 - p_2 \). See Figure 3-1 for an illustration of this stage game:

**Figure 3-1:** An arbitrage opportunity cycle.

Throughout the statement of our results, we will denote the equilibrium choice of any variable \( X \) by \( X^* \).

---

\( ^7 \)Neither the observability of \( X_B \) nor the simultaneity of trades are critical to our qualitative results.
3.3 Equilibrium

3.3.1 Benchmark Equilibria

Prior to determining the equilibria of the game, it is instructive to look at two benchmark outcomes. The first has no arbitrageurs or banks and is referred to as the last-best equilibrium (LB). The second case assumes that arbitrageurs have unlimited wealth and face no financial constraints or, equivalently, that there are no agency problems between the arbitrageurs and the bank. This benchmark is called the first-best equilibrium (FB).

**Proposition 1:** In the last-best equilibrium, the price spread is given by:
\[ \Delta p^{LB} = \lambda \Delta X_N. \]

Meanwhile, in the first-best equilibrium, the arbitrageur’s demand is given by:
\[ X_A^{FB} = \frac{1}{4} \Delta X_N, \]
while the resulting price spread and profit are equal to:
\[ \Delta p^{FB} = \frac{\lambda}{2} \Delta X_N \quad \text{and} \quad \Pi^{FB} = \frac{\lambda}{8} \Delta X_N^2. \]

In the absence of arbitrageurs, the price spread arises because long-term traders absorb \( \Delta X_N \) more units of one asset than the other and require an additional premium of \( \lambda \) per unit of demand imbalance. In the first-best case, since the arbitrageur acts strategically, he eliminates only half of the relative mispricing.

3.3.2 One-Shot Game

We consider subgame perfect equilibria of the one-shot game.

**Proposition 2:** There is a unique SPE and the arbitrageur’s equilibrium strategy is:
\[ R^* = 0 \quad \text{and} \quad X_A^* = \begin{cases} \frac{1}{4} \Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \frac{1}{4} \Delta X_N - \frac{1}{2} X_B & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2} \Delta X_N - 2M_L \\ M_L & \text{otherwise} \end{cases} \]

while the bank’s equilibrium strategy is:
\[ M^* = M_L \quad \text{and} \quad X_B^* = \begin{cases} 0 & \text{if } R = 0 \\ \frac{1}{4} \Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\ \frac{1}{4} \Delta X_N - \frac{1}{2} M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2})M_L \end{cases} \]

Since the bank cannot commit to relax the arbitrageur’s financial constraint or to not copy his strategy, it always sets \( M = M_L \) and \( X_B > 0 \) following communication. Because the bank has price impact, this unambiguously makes the arbitrageur worse off and he prefers not to reveal his information. Arbitrage activity is limited. This also provides more precise foundations for Shleifer-Vishny’s limited arbitrage by identifying an explanation for the separation of brains and capital. In equilibrium, both parties are worse off. By committing to loosen the arbitrageur’s financial constraint and limit its front-running, the bank could achieve positive profits while still making the arbitrageur better off.

Prior to analyzing the repeated game, it is useful to determine the profits to the arbitrageur and the bank in the continuation games from the static case following communication and no communication. The case following communication will determine the bank’s profits when optimally deviating from the relational contract,

\[ \Pi_B^c = \begin{cases} \frac{\lambda}{16} \Delta X_N^2 & \text{if } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\ \frac{\lambda}{8} \Delta X_N^2 - \frac{1}{2} M_L \Delta X_N + \frac{1}{2} M_L^2 & \text{if } \Delta X_N \geq (4 + 2\sqrt{2})M_L \end{cases} \]  

(3.5)

The case without communication will provide the arbitrageur’s outside option:

\[ \Pi_A = \begin{cases} \frac{\lambda}{8} \Delta X_N^2 & \text{if } \Delta X_N \leq 4M_L \\ \lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } \Delta X_N \geq 4M_L \end{cases} \]  

(3.6)

For details, see Lemmas 3 and 4 in the Appendix.

### 3.3.3 Repeated Game

In the infinitely repeated version of the game, we need to make additional assumptions on the information that each arbitrageur has about the past behavior of the bank. Specifically, arbitrageurs know the full history of the bank’s behavior. This degree of knowledge could be rationalized if \( \Delta X_{N,t} \) and \( \Pi_{A,t} \) became known to the market at \( t + 1 \), perhaps because arbitrageurs only remain private for a limited time and the arbitrageur’s terminal payoff is public information. We could also extend the model to incorporate imperfectly observable actions in line with Abreu-Pearce-Stacchetti (1990) and Fudenberg-Levine-Maskin (1994). Nevertheless, this would only amplify our limits to arbitrage problem because the bank would occasionally get away with misbehavior and receive blame when itbehaves.

We also assume that \( \Delta X_{N,t} \) is independent and identically distributed over time. The
independence assumption is reasonable because any predictable component would be known to the entire market and eliminated through competition. The arbitrageur can only expect to profit on the surprise component of $AXN,t$.

Relational contracts are promises the bank makes to the arbitrageur regarding her behavior. However, these promises cannot be enforced by a court. In a given cycle, promises take the form of functions $M^c_t : \mathbb{R} \to [M_L, \infty)$ and $X^c_{B,t} : \mathbb{R} \to [0, \infty)$ that specify the bank’s actions as a function of $AXN,t$. To conserve on notation, we will suppress the dependence of $M^c_t$ and $X^c_{B,t}$ on $AXN,t$. Arbitrageurs cannot make any credible promises to the bank because they only live for one cycle.

We consider efficient subgame perfect equilibria of the repeated game.

**Definition (Efficient Equilibria):** An SPE of the repeated game with payoffs $(\Pi_{1A,t}, \Pi_{1B,t})$ is efficient if and only if there does not exist another SPE of the game with payoffs $(\Pi_{2A,t}, \Pi_{2B,t})$ such that: (i) for every $AXN,t$, $\Pi_{2A,t} \geq \Pi_{1A,t}$, and (ii) $V_{2t} \geq V_{1t}$ where $V_t$ is the value of the relationship to the bank at $t$:

$$V_t = E_t \left[ \sum_{j=1}^{\infty} \delta^{t+j} \Pi_{B,t+j} \right].$$

We discard SPEs that are not efficient because the bank and the arbitrageur can agree to alter their component of the relational contract immediately after $AXN,t$ is realized. Such an agreement is possible if the arbitrageur can be made weakly better off, regardless of initial mispricing, while also improving the bank’s continuation payoff. This is essentially an interim Pareto-optimality criterion (see Brunnermeier (2001)).

In equilibrium, the arbitrageur will only choose $R^*_t = 1$ if the bank can credibly commit to satisfy his individual rationality constraint:

$$\Pi_{A,t}(R_t = 1) \geq \Pi_{A,t}.$$

Similarly, the bank’s choice of $(M^*_t, X^*_B,t)$ can be restricted to those specified by the relational contract, $(M^c_t, X^c_{B,t})$, and the optimal deviation levels, $(M^d_t, X^d_{B,t})$. She chooses to cooperate if and only if:

$$\Pi^c_{B,t} + V^*_t \geq \Pi^d_{B,t}$$

where $V^*_t$ is the continuation value of the financial relationship to the bank in the particular SPE at $t$. Due to the i.i.d. and finite second moment assumptions, this value is bounded. Implicit in the bank’s incentive compatibility constraint is the assumption of maximal punishment by arbitrageurs. Such punishment is credible if there are other equally qualified banks in the market and there is no cost to moving the relationship from one bank to another.

Presumably, this would not be considered a deviation by future generations of arbitrageurs since it does not reflect an action that caused a welfare loss to a previous arbitrageur (i.e. it only produces mutually beneficial gains).
Efficient SPEs induce a structure of optimal collusion between the arbitrageur and the bank. The first element of optimal collusion is illustrated in Lemma 5:

**Lemma 5:** If \( R_t^* = 1 \), then an efficient equilibrium satisfies:

\[
X_{B,t}^* + M_t^* = \frac{1}{4} \Delta X_{N,t}
\]

and

\[
X_{A,t}^* = M_t^*
\]

That is, the arbitrageur and the bank optimally collude to achieve first-best total profits.

When communication occurs, the total demand of the arbitrageur and bank equals that of the first-best equilibrium from Proposition 1. As a result, price spreads are as in the first-best and profits shared between the arbitrageur and bank are maximized. Put simply, if the arbitrageur and the bank are going to collude, they will do so effectively.

Plugging \( X_{A,t}^* \) from the lemma into the arbitrageur’s profit function yields:

\[
\Pi_{A,t}(R_t = 1) = \frac{\lambda}{2} \Delta X_{N,t} \left( \frac{1}{4} \Delta X_{N,t} - X_{B,t}^* \right)
\]

Along with the arbitrageur’s outside option, \( \overline{\Pi}_{A,t} \), we get an upper bound on the bank’s demands:

\[
X_{B,t}^* \leq \overline{X}_{B,t} = \begin{cases} 0, & \text{if } \Delta X_{N,t} \leq 4M_L \\ \frac{1}{4} \Delta X_{N,t} - 2M_L + 4M_L2 \left( \frac{1}{\Delta X_{N,t}} \right), & \text{if } \Delta X_{N,t} \geq 4M_L \end{cases}
\]

This bound is intuitive. If the arbitrageur is unconstrained (i.e., \( \Delta X_{N,t} \leq 4M_L \)), the bank cannot place any demands without making him worse off from communicating with her. When the arbitrageur is constrained, the bank has some freedom to front-run, but only to a limited extent. Likewise, the bank’s IC constraint, along with:

\[
\Pi_{B,t}^c \leq \frac{\lambda}{2} \Delta X_{N,t} \cdot X_{B,t}^*,
\]

implies a lower bound on the bank’s demands:

\[
X_{B,t}^* \geq X_{B,t} = \begin{cases} \max \left\{ 0, \frac{1}{8} \Delta X_{N,t} - \frac{2\overline{V}_t}{\lambda} \left( \frac{1}{\Delta X_{N,t}} \right) \right\}, & \text{if } \Delta X_{N,t} \leq (4 + 2\sqrt{2})M_L \\ \max \left\{ 0, \frac{1}{4} \Delta X_{N,t} - M_L + \left( M_L - \frac{2\overline{V}_t}{\lambda} \right) \left( \frac{1}{\Delta X_{N,t}} \right) \right\}, & \text{if } \Delta X_{N,t} \geq (4 + 2\sqrt{2})M_L \end{cases}
\]

This lower bound indicates that if the bank’s temptation to deviate is high enough, she must
be allowed to trade a certain amount or else she'll choose to deviate.

These two bounds have powerful implications on the possibility of obtaining capital from the bank. Communication between the arbitrageur and the bank is rational if and only if both the arbitrageur's IR and the bank's IC constraints are satisfied (i.e., $\bar{X}_{B,t} \geq \bar{X}_{B,t}$). Proposition 6 shows that this is not always possible.

**Proposition 6:** Information revelation by the arbitrageur cannot be sustained if $\Delta X_{N,t} > \delta x_t^*$ where:

$$
\delta x_t^* = \begin{cases} 
0 & \text{if } V_t^* \leq \lambda M_L 2 \\
8M_L - 4\sqrt{2}M_L 2 - \frac{V_t^*}{\lambda} & \text{if } \lambda M_L 2 \leq V_t^* \leq \left(\frac{1+2\sqrt{2}}{2}\right) \lambda M_L 2 \\
3M_L + \frac{2V_t^*}{\lambda} \left(\frac{1}{M_L}\right) & \text{if } V_t^* \geq \left(\frac{1+2\sqrt{2}}{2}\right) \lambda M_L 2
\end{cases}
$$

This result is illustrated in Figure 3-2:

![Figure 3-2: Illustration of Proposition 6](image)

Figure 3-2: Illustration of Proposition 6
A sketch of the reasoning behind this result can be stated in a straightforward manner. For large enough mispricing (i.e., $\Delta X_{N,t} \geq (4+2\sqrt{2})M_L$), the arbitrageur is constrained even when the bank deviates. Since the bank is unconstrained, she induces a price spread that is half the one that would obtain if the arbitrageur didn’t reveal his information. Therefore, in the event of deviation, the arbitrageur loses exactly half of the profits he would have earned if he chose $R = 0$. This loss is roughly the bank’s gain, and since the arbitrageur’s outside option is unbounded, its value will eventually become greater than the bank’s reputation value $V^*_t$. In other words, stealing about half the arbitrageur’s profits can be better than holding on to future business so the bank cannot commit to improving upon the arbitrageur’s outside option. In this case, communication between the two breaks down and arbitrage is limited like in the one-shot game.

This is the main result of the paper. Financial relationships are limited in their ability to mitigate the hold-up problem between arbitrageurs and banks. More importantly, they are insufficient when mispricing is large so that reputation-based commitment fails when it is needed most for arbitrage. It is important to note that this result cannot be obtained in a standard repeated game with $\Delta X_{N,t}$ held constant across periods. In fact, performing comparative statics on this alternative framework is likely to produce the opposite conclusion that arbitrage is limited when $\Delta X_{N,t}$ is small. To see this, look at the bank’s incentive compatibility constraint under constant mispricing:

$$
\frac{\lambda}{8} \Delta X^2_N - \frac{\lambda}{2} M_L \Delta X_N + \frac{\lambda}{2} M_L^2 \leq \frac{1}{1 - \delta} \left[ \frac{\lambda}{8} \Delta X^2_N - \lambda M_L \Delta X_N + 2\lambda M_L^2 \right].
$$

This reduces to:

$$
\left[ \frac{\lambda \delta}{8} \right] \Delta X^2_N - \left[ \frac{\lambda(1 + \delta)}{2} \right] M_L \Delta X_N + \left[ \frac{\lambda(3 + \delta)}{2} \right] M_L^2 \geq 0.
$$

Though this IC constraint may be violated, it is guaranteed to hold for sufficiently high values of $\Delta X_N$.

The main difference between this setting and ours is that allowing $\Delta X_N$ to be time-varying creates a wedge between the bank’s deviation profit and the value of financial relationships by distinguishing between factors that affect current and future mispricing. This ensures that when current mispricing increases, there is not a corresponding (and amplified) increase in relationship values. As a result, bank commitment becomes more difficult to sustain in times where $\Delta X_{N,t}$ is high.

Some front-running is observed in our equilibrium but its degree is always acceptable to the arbitrageur. Myopically, the bank could do more to hold-up the arbitrageur, but she finds it in her interests to restrain herself. Furthermore, this degree varies with the attractiveness of the arbitrage. The arbitrageur needs to allow the bank to behave more opportunistically when these opportunities are greatest because it is in those events that the temptation to

---

9To simplify exposition, we assume that $\Delta X_N \geq 4(2 + \sqrt{2})M_L$ and that the bank keeps all the surplus from its relationship with the arbitrageur.
deviate is strongest. In this sense, we isolate a phenomenon that is broadly consistent with what Abolafia (Ch. 1, 1996) refers to as "cycles of opportunism" in financial markets.

A second element of optimal collusion that completes our characterization of the efficient equilibria:

**Lemma 7:** An efficient SPE satisfies:

\[
R_t^* = \begin{cases} 
0 & \text{if } \Delta X_{N,t} \leq 4M_L \\
1 & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\
0 & \text{if } \Delta X_{N,t} \geq \delta x_t^*
\end{cases}
\]

This result states that the bank's commitment ability is monotonic. If it can make an adequate commitment at a given level of mispricing, it can also do so at lower levels of mispricing. Lemma 7 also implies that efficient equilibria require the arbitrageur and the bank to collude whenever possible.

We can summarize the characterization of the equilibria to obtain:

**Corollary 8:** A strategy profile is an efficient equilibrium if and only if:

\[
(R_t^*, X_{A,t}^*) = \begin{cases} 
(0, \frac{1}{4} \Delta X_{N,t}) & \text{if } \Delta X_{N,t} \leq 4M_L \\
(1, M_t^*) & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\
(0, M_L) & \text{if } \Delta X_{N,t} \geq \delta x_t^*
\end{cases}
\]

and

\[
(M_t^*, X_{B,t}^*) = \begin{cases} 
(M_L, 0) & \text{if } \Delta X_{N,t} \leq 4M_L \\
\left(\frac{1}{4} \Delta X_{N,t} - X_{B,t}, X_{B,t}\right) & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\
(M_L, 0) & \text{if } \Delta X_{N,t} \geq \delta x_t^*
\end{cases}
\]

where \(X_{B,t} \in [X_{B,t}, \overline{X}_{B,t}]\).

Furthermore, equilibrium price spreads follow immediately from Corollary 8:

**Corollary 9:** Equilibrium price spreads are given by:
\[ \Delta p_i^* = \begin{cases} \frac{\lambda}{2} \Delta X_{N,t} & \text{if } \Delta X_{N,t} \leq \delta x_i^* \\ \lambda \Delta X_{N,t} - 2\lambda M_L & \text{if } \Delta X_{N,t} \geq \delta x_i^* \end{cases} \] (3.8)

This is illustrated in Figure 3-3:

**Figure 3-3:** Equilibrium price spreads
As remarked earlier, whenever there is communication between the arbitrageur and the bank, optimal collusion implies that price spreads are as in the first-best equilibrium. When the channel for informable finance breaks down though, they equal those of the one-shot game and arbitrage only has a fixed effect in correcting mispricing. Figure 3 illustrates the benefits of financial relationships on arbitrage. The thick line denotes the price spread in efficient SPEs of the repeated game. The shaded region represents the gains from relationships. This region becomes larger the more valuable a bank's reputation. An interesting feature of these equilibria is that small differences in initial mispricing, say \( \Delta x^*_t \) versus \( \Delta x^*_{t+1} \), can lead to discontinuous changes in equilibrium mispricing. This is because relationships have their greatest effect in reducing mispricing at \( \Delta x^*_{t+1} = \delta x^*_t + \epsilon \) and no effect on price spreads once communication breaks down.

Under fairly general conditions, the limited arbitrage problem becomes worse as the arbitrageur's initial wealth increases. To simplify the exposition of this result, we make the weak assumption that the surplus allocation rule between the arbitrageur and the bank is independent of \( M_L \).\(^{10}\) In line with the literature on Nash bargaining, surplus is given by the difference between first-best profits and the sum of the arbitrageur's outside option and the bank's net payoff from deviation:

\[
S_t = \Pi_{M,t} - [\Pi_{A,t} + (\Pi_{B,t}^d - V_t)].
\]  

(3.9)

The above assumption implies that there is a process \( \alpha_t \) that does not depend on \( M_L \) and satisfies:

\[
\Pi_{B,t}^d = (\Pi_{B,t}^d - V_t) + \alpha_t \cdot S_t
\]

(3.10)

\[
= \alpha_t \cdot (\Pi_{M,t} - \Pi_{A,t}) + (1 - \alpha_t) \cdot (\Pi_{B,t}^d - V_t).
\]

(3.11)

This result is illustrated in Proposition 10:

**Proposition 10:** Let \( \delta^*_t(X) \) denote the value of \( \delta^*_t \) in an equilibrium with \( M_L = X \). Assuming that \( 0 < M'_L < M''_L \) and \( \delta^*_t(M'_L) > 0 \), it follows that \( \delta^*_t(M'_L) > \delta^*_t(M''_L) \).

Higher arbitrageur wealth has two effects. First, the bank gets lower profits following communication because the arbitrageur's outside option increases. Second, the likelihood that the arbitrageur is constrained and in need of the bank's services decreases. As a result, the value of financial relationships falls when \( M_L \) increases and arbitrage becomes more limited.

### 3.4 Refinement and Examples

Our characterization of efficient equilibria allows for many different possible outcomes because we have not made enough assumptions about how surplus is split between the arbitrageur's bargaining power relative to the bank would increase with \( M_L \). It can be shown that our results in Proposition 10 continue to hold in this case.
trageur and the bank. In this section, we present a simple refinement of our equilibrium, based on competition between banks in forming relationships with arbitrageurs, that pins down this allocation of surplus and narrows our set of equilibria down to a unique one.

3.4.1 Simple Refinement

Consider an extension of our stage game that allows banks to make offers of relational contracts to the arbitrageur after he receives his private signal. Such promises prescribe a contingent action plan for the bank as a function of $\Delta X_{N,t}$. This implicitly defines a surplus allocation rule between the bank and the arbitrageur. We consider two extreme cases of competition between banks. In the first case, which we refer to as monopoly, the bank is alone in making an offer. In the second case, there are a large number of banks who induce perfect competition in bidding.

In the monopoly case, the bank gets all the surplus from the relationship. As a result, if the arbitrageur and the bank cooperate, the latter receives the first-best level of profits minus the arbitrageur’s outside option. This implies a unique efficient SPE that is characterized by $V_t^*$. This relationship value is given by the largest solution to the fixed-point problem:

$$V_t^* = \frac{\delta}{1 - \delta} \int_0^{\delta t_*} \Pi_{B,t}^{c,M}(\Delta X_{N,t})dF(\Delta X_{N,t})$$

where $\Pi_{B,t}^{c,M}$ is defined in Lemma 11 in the Appendix. Existence of the unique solution, $V_t^*$, is guaranteed since zero is always a solution to the equation. Further, $V_t^*$ is independent of time so the resulting equilibrium is stationary.

Under perfect competition, the bank promises the arbitrageur as much surplus as she possibly can. In other words, she either sets $X_{B,t}^* = 0$ or, when she does trade, binds herself to her IC constraint by setting:

$$\Pi_{B,t}^{c,PC} + V_t^* = \Pi_{B,t}^{d}.$$  

As in the monopoly case, there is a unique efficient SPE characterized by $V_t^*$ which is the largest solution to the problem:

$$V_t^* = \frac{\delta}{1 - \delta} \int_0^{\delta t_*^{PC}} \Pi_{B,t}^{c,PC}(\Delta X_{N,t})dF(\Delta X_{N,t}).$$

---

11With the exception of Proposition 10, we’ve allowed $X_{B,t}^*$ to be any curve in the shaded area of Figure 2 (i.e. anything between $X_{B,t}^*$ and $\bar{X}_{B,t}$).

12Using the term monopoly here is a bit abusive since we’ve motivated some earlier results with the assumption that the arbitrageur can start up new relationships with other equally proficient banks.

13It is easy to verify that, if there are two solutions to the fixed-point problem with values $V_t^*$ and $V_t^{**} > V_t^*$, moving from the equilibrium implied by $V_t^*$ to $V_t^{**}$ is an interim Pareto improvement. Therefore, only the equilibrium with $V_t^{**}$ is an efficient SPE.

112
where $\Pi_{B,i}^{PC}$ is defined in Lemma 12 in the Appendix. This produces a solution, $V_{PC}^*$, that is also independent of time.

In both environments, stationarity is driven by the independence assumption on $\Delta X_{N,i}$ and the fixed surplus allocation rule. It also requires that arbitrageurs be short-lived. This is a different foundation than found elsewhere in the relational contracting literature. For example, the weak stationarity lemma in Levin (2003) is driven by the unlimited wealth of the principal and agent which allows for settling up on a period-by-period basis. Nevertheless, it should be noted that stationarity of $V$ is not important for the main results of this chapter.

Communication is easier to sustain in the monopoly case than the perfect competition case, since aggressive competition between banks forces them to bid away their share of the surplus and lowers their valuation of financial relationships. This worsens their ability to make commitments. Proposition 13 formalizes this intuition:

**Proposition 13:** Comparing the efficient SPE from the monopoly and perfect competition cases, we have that:

$$V_M^* \geq V_{PC}^*$$

and

$$\delta x_M^* \geq \delta x_{PC}^*.$$

### 3.4.2 Binomial Distribution Example

Assume that $\Delta X_{N,i}$ equals $\epsilon > 0$ with probability $p$ and 0 otherwise. Let $M_L = \beta \epsilon / 4$ with $\beta < 1$ so that the arbitrageur is constrained when there is mispricing. The smaller the value of $\beta$, the more severely constrained is the arbitrageur when $\Delta X_{N,i} = \epsilon$.

When the bank has monopoly power, we can solve for the efficient equilibrium by assuming that information revelation can be sustained at $\Delta X_{N,i} = \epsilon$ and checking for consistency afterwards. If there is communication, we have:

$$V_M^* = \left( \frac{\delta}{1 - \delta} \right) E \left[ \Pi_{FB} - \Pi_{A} \right] = \frac{\lambda p}{8} \left( \frac{\delta}{1 - \delta} \right) (1 - \beta) 2\epsilon 2.$$

We check for consistency by verifying whether the bank’s IC constraint is satisfied. It is if:

$$\beta \leq \bar{\beta}_M = \frac{4\theta + 2 - 2\sqrt{\theta + 1}}{4\theta + 3}$$

where $\theta = p\delta / (1 - \delta)$. If the IC constraint doesn’t hold, $V_M^*$ equals zero, since no relational contract is enforceable.

Notice that $\bar{\beta}_M$ is monotonically increasing in $\theta$. This implies that there are no limits to arbitrage if mispricing is sufficiently frequent or the bank is patient enough. This is intuitive because both lead to higher relationship values for the bank. What may be surprising is that
and $\epsilon$ do not affect $\tilde{\beta}_M$. This is due to scaling effects: all profit functions are perfectly linear in $\lambda$ and quadratic in $\epsilon$, so the arbitrageur’s IR and the bank’s IC constraints are unaffected by these parameter value. In the case of $\epsilon$, the scaling argument requires us to hold $\beta$ fixed because of the arbitrageur’s outside option.

Likewise, in the case of perfect competition, if communication between the arbitrageur and the bank occurs at $\Delta X_{N,t} = \epsilon$, we have:

$$V_{PC}^* = \frac{\lambda \theta (\frac{1}{8} - \frac{\theta}{8} + \frac{\theta^2}{32}) \epsilon^2}{1 + \theta}. \tag{3.12}$$

This can be sustained if the arbitrageur’s IR constraint holds. That is, if:

$$\beta \leq \tilde{\beta}_{PC} = \frac{4\theta + 2 - 2\sqrt{\theta + 1}}{4\theta + 3} \tag{3.13}$$

The same results and intuitions from the monopoly case hold here. Surprisingly, we have that $\tilde{\beta}_{PC} = \tilde{\beta}_M$ in this setting. This is a particular feature of the binomial distribution example. When there is only one level of possible mispricing, competition does not matter at the margin where arbitrage becomes limited because there is no surplus, beyond the bank’s required rent from her IC constraint, to bargain over. To investigate the effect of competition between banks on the limits to arbitrage problem requires multiple potential levels of mispricing in order to produce regions where the division of surplus matters. This motivates our next example.

### 3.4.3 Normal Distribution Example

We assume that $X_{N_1,t}$ and $X_{N_2,t}$ are joint normally distributed so that we can write $f(\Delta X_{N,t}) = 2 \cdot n(\Delta X_{N,t}; \sigma^2)$ where $n(\cdot; \sigma^2)$ is the density of a normal distribution with mean 0 and standard $\sigma$. We also set $\beta = 4M_L/\sigma$. This parameter can be interpreted as the standard deviation shock to $\Delta X_{N,t}$ that is required to make the arbitrageur constrained under his default financial constraint. This is economically similar to the $\beta$ parameter from the binomial distribution case. We also write $\langle \delta x^* \rangle \equiv \delta x^*/\sigma$. This is the threshold where arbitrage becomes limited normalized by the standard deviation of mispricing. Holding $\langle \delta x^* \rangle$ constant fixes the frequency of limited arbitrage in this setting.

As in the binomial case, the threshold point is invariant to scaling, as shown in the following Lemma:

**Lemma 14:** In both the monopoly and perfect competition cases, holding $\beta$ constant, $\langle \delta x^* \rangle$ does not depend on $\sigma$ and $\lambda$.

This example cannot be solved in closed form, so we provide numerical solutions. It is useful to define $c = 2 \cdot N(-\beta)$ where $N(.)$ is the cumulative distribution for a standard normal. This value represents the percentage of arbitrage opportunities where the arbitrageur
is constrained under his default financial constraint. Figure 3-4 illustrates the value of $\langle \delta x^* \rangle$ as a function of $c$ and the discount factor $\delta$ in both the monopoly and perfect competition cases:

![Diagram](image)

**Figure 3-4: Relationship thresholds**

Notice that there is a substantial difference in the thresholds between the perfect competition and monopoly cases. Consistent with Proposition 13, we observe limited arbitrage more frequently in the case of perfect competition among banks. For instance, as illustrated in the figure, when $\delta = 0.85$ and the arbitrageur is constrained in 62.5% of the arbitrage opportunity cycles, limited arbitrage is over 40,000 times more likely in the case of perfect competition ($2 \cdot N(-2.88)$ vs. $2 \cdot N(-5.33)$). The frequency of limited arbitrage is also economically significant. If the arbitrageur is sufficiently wealthy (low $c$), then financial relationships have no value and outcomes are as in the one-shot game. This is expected since the arbitrageur rarely needs the bank and has a high outside option, thus producing a low value to reputation that the bank is unable to ever meet the arbitrageur’s outside option.

Perhaps surprisingly, Figure 3-4 also indicates that a small change in $c$ can have a large and discontinuous impact on the role that relationships play in funding arbitrage activity. This is due to the arbitrageur’s outside option. As he becomes wealthier, he requires a larger share of profits, which lowers the value of relationships to the bank. This reduction in value also lowers $\delta x^*$, which further lowers $V^*$. The discontinuity arises, because $V^* \geq \lambda M_L2$ is needed to satisfy the bank’s IC constraint at $\Delta X_{N,t} = 4M_L$. 

115
3.5 Additional Extensions and Discussion

3.5.1 Explicit Contracting

In this section, we consider a special case of explicit contracting, namely the possibility of signing labor contracts with banks. A labor contract specifies a non-negative and non-decreasing wage payment that is conditional on profits and possibly a trading budget that sets an upper bound on the position, $\bar{M}$, the bank employee can undertake. However, the content of communication between the employee and the bank is still non-contractible. Employment contracts can be renegotiated once an employee joins the bank and the bank maintains her ability to attract employees even if she deviates from her implicit agreements with arbitrageurs who trade on their own.

There are two types of risk-neutral agents: arbitrageurs, identical to the ones from previous sections, and speculators. A proportion $\theta$ of potential employees are speculators and the cumulative distribution function of $\Delta X_{N,t}$ is monotonically increasing. Speculators have no private information and can only make risky investments that generate mean zero profits with distribution $G$ per unit. Due to their limited liability constraint, they may extract rents from the bank by risk-shifting. Speculators also have access to an alternative employment opportunity that pays a fixed wage $\bar{w} > 0$. Arbitrageurs’ outside opportunity is to trade on their own. For simplicity, we will assume that the bank is a monopolist and $\theta$ is sufficiently close to 1.\(^\text{14}\)

The latter induces a fly-by-night constraint where the bank chooses to screen out all speculators. The optimal contract between the bank and its employees is given in Proposition 15:

**Proposition 15:** Let $\Omega$ denote the set of arbitrageur types $\Delta X_{N,t}$ that are hired by the bank. We have that: (i) the optimal employment contract specifies a finite upper bound $\bar{M}$, (ii) $\Omega \subseteq [M_L/4, \bar{M}/4]$, (iii) the speculators’ screening condition binds:

$$\bar{w} = \max_{X \leq \bar{M}} E^G [W(X \cdot \tilde{\Pi})], \quad (3.14)$$

(iv) the wage offer is given by:

$$W(\Pi) = \begin{cases} 
0 & \text{if } \Pi \leq 2\lambda M_L^2 \\
2M_L\sqrt{2\lambda\Pi} - 2\lambda M_L^2 & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda \bar{M}^2 \text{ and } 2\sqrt{2\Pi} / X \in \Omega \\
\sup_{\Pi < \Pi} W(\Pi) & \text{if } 2\lambda M_L^2 \leq \Pi \leq 2\lambda \bar{M}^2 \text{ and } 2\sqrt{2\Pi} / X \notin \Omega \\
4\lambda M_L\bar{M} - 2\lambda M_L^2 & \text{if } \Pi \geq 2\lambda \bar{M}^2
\end{cases}, \quad (3.15)$$

\(^{14}\)This assumption is made for expositional reasons and the results continue to hold if $\theta \in (0, 1)$. 

116
and (v) Arbitrage is still limited.

The main parts of this proposition are (i) and (v). Both follow from the fact that the fly-by-night constraint induces the bank to limit the rents speculators can extract. In order to do this, the bank cannot leave both the maximal wage and trading budget in its employment offer unbounded. The upper bound on wages, which is a function of $M$, follows from the requirement that the contract be renegotiation proof. Unfortunately, these bounds also screen out the arbitrageurs with the most profitable arbitrage opportunities. Hence, arbitrage activity is still limited. It should be noted that the form of contract obtained here is similar to those observed in practice. For instance, proprietary traders employed by investment banks are often subject to position limits and receive bonuses that depend on their trading profits.

Interestingly, allowing for explicit contracts can increase the frequency of limited arbitrage. This is due to the fact that, under explicit contracting, the bank has less to lose from deviation because it keeps its profits from continuing to employ proprietary traders.

**Proposition 16:** If $\delta x^*_M > 4\bar{M} > 4M_L$, then allowing explicit contracts leads to more severe limits to arbitrage.

The observation that explicit contracts can crowd out implicit ones is also emphasized, in a managerial compensation setting, by Baker-Gibbons-Murphy (1994).

### 3.5.2 Risky Arbitrage

So far, we have assumed that an arbitrage opportunity is riskless. It turns out that risky arbitrage can only worsen the limits to arbitrage problem because it can lead to the arbitrageur becoming financially distressed and being forced to liquidate his positions when mispricing worsens. In these states of the world, the value of financial relationships is lower since the arbitrageur is less likely to survive. Larger mispricing also increases the bank's temptation to deviate, which amplifies the decline in relationship value. Therefore, even if the arbitrageur is not currently worried about expropriation, he might be in the future. This induces both delay in communication and excessive restraint on the part of the arbitrageur. A thorough analysis of equilibrium in this extended environment is beyond the scope of this chapter.

### 3.5.3 The Market Structure of Arbitrage

In addition to investment banks, several other organizations, like fund-of-funds and seeders, have emerged in the market for funding arbitrageurs. Fund-of-funds invest in a portfolio of established hedge funds while seeders help emerging managers obtain initial capital to get up and running. These investors have significantly more expertise than most investors and usually negotiate favorable liquidity treatment from the funds they invest in. Therefore,
they can play the role of informable financier for broad strategy at a fund’s initiation as well as temporary investment opportunities that arise during the course of operations.

Furthermore, without their own trading desks, these organizations cannot directly implement the arbitrageur’s strategy. They can only indirectly expropriate by sharing information about the arbitrageur’s profit opportunities with other funds they invest in. This is similar to the hold-up problem investigated by Cestone-White (2003). Interestingly, consistent with the optimal contracting solution derived to overcome this problem, fund-of-funds and seeders often acquire equity-like claims in the hedge fund’s management company.

The existence of these institutions is consistent with certain implications of our theory. Namely, that there are benefits to setting up institutions that can serve as informable financiers without having a substantial ability to expropriate the arbitrageur. However, it must be noted that investment banks still provide the bulk of financing to arbitrageurs.

There are several potential explanations for this fact. The industry may still be maturing and facing its own wealth constraints. Alternatively, renegotiation of margin agreements and the reallocation of margin use across strategies may be less costly and time consuming than obtaining additional finance from fund-of-funds and seeders. This would explain why investment banks are especially dominant as a source of short-term financing. The industry may also be less effective in understanding and monitoring the arbitrageurs’ strategies because of its lack of direct participation in trading.

Regarding the organizational structure of banks, there is a debate on whether Chinese walls should be required between prime brokerage and proprietary trading desks. Our model illustrates one benefit of the absence of Chinese walls, namely that communication between divisions leads to improved arbitrage when prime brokers lack the expertise to understand the strategies themselves. However, these walls will be useful whenever the prime broker can understand the arbitrageur because the expropriation threat is absent when he is prohibited from leaking information to proprietary traders. Therefore, the optimality of enforcing Chinese walls between the two divisions depends on a trade-off between these factors.

3.5.4 Financing Innovation

Our model can also be applied to the more general setting of financing innovation. Consider a setting where a sequence of entrepreneurs want to fund projects with public debt and private equity. This is isomorphic to our model in sections 2 through 4, where $\Delta X_{N,t}$ now represents the profitability of a project.¹⁵ Let $M_L$ denote the production capacity that the entrepreneur can achieve while relying only on arms-length debt financing. The entrepreneur can also obtain capital through a relationship with an informable financier, as in our previous model. Likewise, the extension of section 5.1 can be viewed as allowing for other types of securities, like equity, to be issued when raising capital without disclosing private information.

An interpretation of the informable financier in this context would be a venture capitalist, who has expertise in specific industries and can gauge the quality of a project, but also has the contacts to implement the project herself. As suggested in the quote below, reputation

¹⁵This can explicitly be viewed as a consumer demand shock in the spirit of Rotemberg-Saloner (1986).
is an important concern here:

*We are extremely conscious that corporations and entrepreneurial ventures can be in conflict... if [the entrepreneurs] don't know us or have never interacted with Cisco, there is an initial concern that needs to be overcome... We’ve overcome such concerns by building a track record. We have enough references within the venture capital community that we can say, ‘Hey, why don’t you talk to John Doerr or go to Don Valentine and ask them what they think about having Cisco as an investor.’ Pretty unanimously we get over the hurdle.*

- Mike Volpi, Head of Cisco Ventures (in Gupta p.32)

However, our model suggests that reputation alone cannot always guarantee first-best implementation. The most profitable projects will still be undercapitalized.

### 3.6 Conclusion

We provide an economic foundation for the limits to arbitrage problem, based on a hold-up problem between arbitrageurs and informable financiers. Reputational concerns on the part of the financiers partially mitigate this problem. However, relationships fail when needed most. Additionally, more bargaining power in the hands of arbitrageurs and increased competition among banks can make everyone worse off since it reduces the value of relationships to the bank. Holding the degree of bargaining power fixed, we also show that higher initial arbitrageur wealth worsens the effectiveness of informable finance. Finally, allowing for explicit contracts does not necessarily alleviate the limits to arbitrage problem and, in fact, can worsen it.
Appendix A

Chapter 1: Appendix

A.0.1 A model with two capital stocks

The model presented in section 3 is designed to capture the basic intuition of a reallocation shock, and it holds the capital in the investment sector fixed. In this section I relax the assumption that the capital stock in the investment sector is fixed, whereas the other elements of the model remain the same. I find that the basic insights from the baseline model are unaltered.

The investment goods sector (the I-sector) produces the investment good using sector specific capital $K_I$ and labor $L_I$. The output of the I-sector can now be used to increase the capital stock in either of the two sectors according to

$$ c \left( \frac{I_{I,t}}{K_{I,t}} \right) K_{I,t} + c \left( \frac{I_{C,t}}{K_{C,t}} \right) K_{C,t} \leq a Y_t K_{I,t}^{\beta_I} L_{I,t}^{1-\beta_I}, \quad (A.1) $$

where $c(\cdot)$ is an increasing and weakly convex function. As in the baseline model, the shock $Y$ represents a shock to the productivity of the I-sector. A positive shock to $Y$ increases the output of the investment good holding capital and labor fixed and implies that the economy has to give up fewer units of the consumption good in order to increase the capital stock in either sector. Firms in the I-sector have a Cobb-Douglas technology, where $\beta_I$ is the elasticity of output with respect to capital. Firms face constant returns to scale with respect to capital because labor is flexible. However, because the supply of labor is fixed at the aggregate level, the investment sector is subject to decreasing returns to scale. Finally, $\alpha$ is a parameter controlling the relative size of the two sectors. Capital allocated in the investment sector depreciates at a rate $\delta$, while investment in the I-sector is denoted by $I_I$.

Investment in both sectors is irreversible

$$ I_C \geq 0, \quad (A.2) $$

$$ I_I \geq 0. \quad (A.3) $$

Equations (A.2), (A.3) and (A.1) imply that there is an upper bound and a lower bound on investment in each sector. The lower bound comes from the irreversibility constraints,
whereas the upper bound stems from the fact that total investment cannot exceed the output of the investment sector. If \( c(\cdot) \) satisfies Inada conditions, these constraints will not be binding however.

The capital stock in the I-sector evolves according to

\[
dK_{I,t} = I_{I,t} dt - \delta K_{I,t} dt,
\]

and the investment shock follows (1.6). The value of a representative firm in the investment sector equals

\[
S_{I,t} = \int_t^{\infty} \frac{\pi_s}{\pi_t} \left( \lambda_s Y_s K_{I,s}^{1-\beta_I} - w_s L_{I,s} - \lambda_s c \left( \frac{I_{I,s}}{K_{I,s}} \right) K_{I,s} \right) ds,
\]

where \( w_t \) is the wage and \( \lambda_s \) is the price of the investment good in terms of the consumption good at time \( t \).

**Proposition 3.** The social planner's value function is

\[
J(X, Y, K_I, K_C) = \frac{(X K_C^{\beta_C})^{1-\gamma}}{1-\gamma} f(z, \psi),
\]

where \( f(z, \psi) \) satisfies the PDE in the appendix and

\[
\psi_t \equiv \ln(K_{C,t}) - \ln(K_{I,t})
\]

and

\[
z_t \equiv \ln(Y_t) + (\beta_I - 1) \ln(K_{I,t}).
\]

The optimal investment policies, \( i^*_C \) and \( i^*_I \), satisfy

\[
((\beta_I - 1)f_x(z, \psi) - f_\psi(z, \psi)) c^{-1}(a e^z L_{I,t}^{1-\beta_I} - c(i^*_C, t) e^\psi) c'(i^*_C, t) e^\psi
\]

\[
= (\beta_C(1-\gamma)f(z, \psi) + f_\psi(z, \psi))
\]

\[
i_{I^*, t} = c^{-1}(a e^z L_{I,t}^{1-\beta_I} - c(i^*_I, t) e^\psi)
\]

and the optimal allocation of labor satisfies \( L^*_I = 1 \) and \( L^*_C = 1 - l \), where \( l(z, \psi) \) satisfies

\[
Q(z, \psi) l(z, \psi)^{-\beta_I} e^z = \frac{1 - \beta_C}{1 - \beta_I} (1 - l(z, \psi))^{-\gamma(1-\beta_C)-\beta_C},
\]

and

\[
Q(z, \psi) \equiv \left( \frac{(\beta_I - 1)f_x(z, \psi) - f_\psi(z, \psi)}{1-\gamma} \right) \frac{1}{c'(i^*_I)}.
\]

In equilibrium, the optimal policies now depend on two state variables. The first state variable, \( z \), captures the marginal benefit of an additional unit of capital invested in the
I-sector. ¹ The second state variable equals the ratio of the capital stocks between the two sectors. As long as the parameter restrictions in the Appendix are satisfied, the state vector has a unique stationary distribution. This guarantees that in equilibrium, both capital stocks are used and that one sector does not dominate the economy. Innovations to z come only from the I-shock (Y).

The concavity of J implies that the investment policy has a bang-bang property. If ψ is below ψ(z), all investment goes to the I-sector, pushing the ratio up towards the target, and vice versa. This policy results from the investment constraints. Ideally, the planner would like to set the ratio of effective capital stocks in both sectors equal to some function of z. However this is not possible because z has a diffusion component while the planner has limited ability to control the capital stock. Instead, the planner has a target capital stock in mind and imperfectly controls ψ towards that target. The social planner invests in the I-sector if the capital stock in the I-sector is too low relative to the capital stock in the consumption sector (ψ is low) or if the productivity of capital in the I-sector (z) is high. Figure A-2(b) plots the investment threshold ψ(z).

The relative price of the investment good is

$$\lambda = \frac{J_{K_C}}{U_C} \frac{1}{c'(s_C^*)}.$$ 

The relative price of the investment good equals the marginal value of capital in the unconstrained sector divided by marginal installation cost and marginal utility.

The optimal policy for l is similar to that in the simplified model. It is increasing in Y as long as λ(Y)Y is increasing in Y, which is generally the case as long as the inverse demand curve for the investment good is sufficiently inelastic. Enabling the economy to invest in the I-sector does not alter the fact that an I-shock carries a negative premium. There is an additional decision to be made, namely in which sector will the investment good be used to increase the capital stock, but the labor allocation decision is essentially unaltered. In figure A-2(d), I plot the labor allocation policy l(z, ψ).

The relative price of investment versus consumption firms is plotted in figure A-3(e). As in the baseline model, the relative price of the two sectors is increasing in the state of investment-specific technology, z. Consequently, the IMC portfolio is positively correlated with the investment-specific shock. Finally, one can compute the value of assets in place and growth opportunities for the whole economy, as in section 3.5.2. The value of assets in place in the consumption S_C,t and investment sector S_I,t are

$$S_{C,t}^Y = \max_{L_{C,s}} E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( X_s(K_{Ct}e^{-\delta(t-s)})^{\beta_C} L_{C,s}^{1-\beta_C} - w_s L_{C,s} \right) ds.$$  

¹Because the I-sector faces decreasing returns to scale, the capital stock in the I-sector becomes a state variable. If instead the I-sector had constant returns to capital on the aggregate, optimal policies would depend only on one state variable which would be equal to ln Y + ln K_I - ln K_C.
and
\[
S_{I,t}^V = \max_{L_t} E_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( aY_t(K_{I,t}e^{-\delta(s-t)}\beta_t \hat{L}_{I,t}^{1-\beta_t} - w_t \hat{L}_{I,t}) \right) ds. \tag{A.7}
\]

The relative value of assets in place over future growth opportunities for the whole economy equals
\[
\frac{S_Y^V}{S_G^V} = \frac{S_{I,t}^V + S_{C,t}^V}{S_{M}^V - S_{I,t}^V - S_{C,t}^V},
\]
and is computed in the proof of Proposition 3 in the Appendix. Figure A-3(f) plots the relative value of assets in place to future growth opportunities. As in the baseline model, the ratio is decreasing in \( z \). Positive shocks to investment-technology lower the cost of new investment, increasing the value of future growth opportunities relative to assets in place.

### A.0.2 Proofs and Derivations

**Proof of Proposition 1.** First let’s establish that the value function is bounded. To find a lower bound we need to compute a feasible policy and show that it provides the agent with finite utility. We shall prove this for the CRRA case. Consider the policy \( L_{I,t} = 0 \). Such a policy is clearly feasible, and it gives the agent:

\[
J_{LB}^* = E_0 \int_0^\infty \frac{(K_0^\beta X_0)^{1-\gamma}}{1-\gamma} e^{-\rho t - (1-\gamma)\delta C t} \mu x t^{-1/2(1-\gamma)\sigma_x^2 + (1-\gamma)\sigma_x (Z_t - Z_0)} dt.
\]

So, if:
\[
u \equiv \rho - (1-\gamma)(\mu x - \beta \delta C) + \frac{1}{2}\gamma(1-\gamma)\sigma_x^2 > 0
\]

then
\[
J_{LB}^*(K_0, X_0) = \frac{(K_0^\beta X_0)^{1-\gamma}}{(1-\gamma)(\rho - (1-\gamma)(\mu x - \delta C) + \frac{1}{2}\gamma(1-\gamma)\sigma_x^2)}
\]
and \( J_{LB}^* = -\infty \) otherwise.

Incidentally, this also characterizes the behavior of \( f(\omega) \) at \( \omega \to -\infty \). So given this:
\[
J(K_0, X_0, \omega, \psi) \geq J_{LB}^*(K_0, X_0) \Rightarrow f(\omega) \leq \frac{1}{u}
\]

Deriving an upper bound in the case of \( \gamma > 1 \) is trivial since utility is bounded above by zero.
The Hamilton-Jacobi-Bellman equation for the social planner's optimization problem is:

\[
0 = \max_{L_t, L_c, i_t, i_c} \left\{ U(C) + (i_C - \delta_C)J_{KC}K_C + J_XX\mu_X + \frac{1}{2}J_{XX}X^2\sigma_X^2 + +\mu_YJ_YY + \frac{1}{2}J_{YY}Y^2\sigma_Y^2 - \rho J \right\}
\]

subject to:

\[
\begin{align*}
C &\leq XK_C \beta L_C^{1-\beta C} \\
C(i_C)K_C &\leq \alpha Y K_I^\beta L_I^{1-\beta I} \\
i_C &\geq 0 \\
L_C + L_I &\leq 1
\end{align*}
\]

The planner's problem can also be written in sequence form

\[
L(X_0, Y_0, K_{C,0}, K_{I,0}; i_t, i_c) = \inf_{\lambda, \pi, \lambda^2, \lambda^3} \sup_{i_t, i_c, L_t, L_c} E_0 \int_0^\infty e^{-\rho s} U(C_s) - \\
\lambda_s \pi_s \left( i_C K_C - \alpha Y K_I^\beta L_I^{1-\beta I} \right) \\
- \pi_s (C_s - X_s K_C \beta L_C^{1-\beta C}) - w_s (L_t, L_c, 1) \\
- \lambda^2 i_C - \lambda^3 t ds 
\]

An application of the Feynman-Kac theorem on \( L \) and using the envelope theorem yields:

\[
0 = \inf_{\lambda, \pi, \lambda^2, \lambda^3} \max_{t \in (0,1), i_C} \left\{ U(C) + J_{KC} K_C (i_C - \delta) + \mu_X J_X + \frac{1}{2}J_{XX}X^2 + \mu_Y J_Y Y + \frac{1}{2}J_{YY}Y^2 - \rho J - \lambda \tilde{\pi} \left( i_C K_C - \alpha Y K_I^\beta L_I^{1-\beta I} \right) - \lambda^2 i_C - \tilde{\pi}(C - X K_C \beta (1 - l)^{1-\beta C}) \right\}
\]

The first order condition with respect to investment also implies that

\[
\lambda \pi' (i_C^*) = J_{KC}
\]

Where \( \lambda \pi \) is the lagrange multiplier for the investment resource constraint.

The first order conditions with respect to \( L_C \) and \( L_I \) along with the resource constraint
on labor:
\[(1 - \beta_C) \left( X K^\beta_C (1 - L_I)^{1-\beta_C} \right)^{1-\gamma} (1 - L_I)^{-1} = \alpha (1 - \beta_I) J_{K_C} Y K^\beta_I L_I^{-\beta_I} \times c^{-1} (\alpha Y K^\beta_I L_I^{1-\beta_I}) \]

We will look for a guess of the form:
\[J = \frac{(X K^\beta_C)^{1-\gamma}}{1-\gamma} f(\omega)\]

where \(\omega \equiv \ln \frac{Y}{K_C}\). Writing \(l = L_I\), the first order condition for \(l\) is:
\[(1 - l(\omega))^{(1 - \beta_C)(1 - \gamma) - 1} = \frac{\alpha (1 - \beta_I)}{1 - \beta_C} \left[ \epsilon^\omega \left( \beta_C f(w) - \frac{f_\omega(w)}{1-\gamma} \right) \right] c^{-1} (\alpha e^\omega l(\omega)^{1-\beta_I}) l(\omega)^{-\beta_I} \]

Whereas the HJB equation reduces to an ODE in \(f\):
\[0 = \left\{ (1 - l)^{(1 - \beta_C)(1 - \gamma) + c^{-1} (\alpha e^\omega l^{1-\beta_I}) (\beta_C (1 - \gamma) f - f')} \right\} \]
\[-u f + (\mu_Y + \delta_C) f' + \frac{1}{2} \sigma_y^2 (f'' - f') \]

This reduces to a DAE system, that unfortunately does not admit an analytic solution. Instead, I solve the optimization problem numerically, as described in Section B.

A competitive equilibrium can be constructed from the solution above. Consider the following candidate, where \(\lambda^* = \lambda\), \(w^* = w\), \(L_C^* = 1 - l(\omega)\), \(L_I^* = L_I(\omega)\), \(I_C^* = I_C\), \(\pi^* = \pi^\star\) I need to show that this satisfies the conditions in the definition of the competitive equilibrium. Without loss of generality, assume that there is only one competitive firm in each sector. The firm in the consumption sector purchases capital goods at an absolute price \(\pi^* X^C \equiv X^C\), hires labor at an absolute price \(\pi^* w^*\) and sells its output at price \(\pi^*\). It chooses \(i^*_C\) to maximize its value
\[\pi^* X^C s\pi^* s^* = E_t \pi^* X^C sK^\beta_C L_C^{1-\beta_C} - c(i^*_C) K^\beta_C \pi^* s^* - L_C \pi^* s^* \]

Its HJB equation can be written as
\[0 = \max_{i_C, L_C} \left\{ \pi^* X^C K^\beta_C L_C^{1-\beta_C} - c(i_C) K^\beta_C \pi^* s^* - L_C \pi^* s^* + \tilde{S}_{K_C} K^\beta_C (i_C - \delta) + D_{X,Y} \tilde{S} \right\} \]

A firm’s FOC implies \(\pi^* \lambda^* c'(i^*_C) = \tilde{S}_{K_C}\), whereas the proof of proposition 2 below implies that \(\tilde{S}_{K_C} = J_{K_C}\). Similarly, it picks labor such that:

\[X K^\beta_C L_C^{1-\beta_C} = w^*\]
The firm in the investment sector on the other hand sells its output at price $\lambda^*$ and thus picks $L^*_I$ such that

$$a\lambda^* Y_t K^\beta_I L^*_I = w^*$$

Labor market clearing implies that $L^*_C + L^*_I = 1$ and combining this with the last two equations gives equation (A.8) so the social planner’s optimal labor choice coincides with firm optimal policies given equilibrium wage $w^*$ and the labor market clears. Now, we have shown that at prices $\lambda^*$ and $w^*$ 1) firms pick the same policies as the social planner and 2) labor and capital goods market clears. Now a household with total wealth $W_t$ picks consumption policy $C^*$ to maximize its lifetime utility

$$\tilde{J} = E_t \int_t^\infty e^{-\rho_s} U(C_s) \, ds$$

subject to $W_t = E_t \int_t^\infty C_s \pi_s \, ds$. Its total wealth equals its financial wealth plus is human capital, so

$$W_t = E_t \int_t^\infty w_s ds + S_{C,t} + S_{I,t}$$

Because markets are dynamically complete, it purchases consumption at a price (relative to probability) $\pi^*_s$. Given $\pi^*_s$, the household is happy to follow the equilibrium policy, since

$$\pi^*_s = e^{-\rho_s} U'(C^*_s)$$

and the equilibrium consumption policy clears the consumption goods market by construction. \qed

**Proof of Proposition 2.** Consider the value of a firm in the C-Sector. The firm buys new capital and hires labor to maximize its value

$$\pi_t S^C_t = E_t \int_t^\infty \pi_s \left(X_s K^\beta_{C,s} L^*_C - w_s L_{C,s} - \lambda_s c(iC) K_C \right)$$

the intratemporal first order condition with respect to labor imply that:

$$(1 - \beta_C) X_s K^\beta_{C,s} (L^*_C)^{1-\beta_C} = w_s L^*_C$$

so firm value is:

$$\pi_t S^C_t = E_t \int_t^\infty \pi_s \left(\beta_C X_s K^\beta_{C,s} (L^*_C)^{1-\beta_C} - \lambda_s c(iC) K_C \right) ds$$

The planner’s lagrangian evaluated at the optimum can be written as:

$$\mathcal{L} = E_0 \int_0^\infty e^{-\rho_s} U(C_s) - \pi_s (C_s - X_s K^\beta_{C,s} L^*_C^{1-\beta_C}) - \lambda_s \pi_s \left(c(iC) K_C - aY_s K^\beta_I L^{1-\beta_I} \right) ds$$

127
The envelope theorem implies that

\[ \frac{\partial L}{\partial K_c} = \frac{\partial J}{\partial K_c} \]

also note that

\[ \frac{\partial K_{C,s}}{\partial K_{C,0}} K_{C,0} = K_{C,s} \]

Therefore

\[ \frac{\partial J}{\partial K_{C,0}} K_{C,0} = E_t \int_t^\infty \pi_s \left( \beta_C X_s K_{C,s}^{\beta_C} (L_{C,s}^*)^{1-\beta_C} - \lambda_s c(i_{C,s}) K_{C,s} \right) \, ds \]

Similarly, the value of a firm in the investment sector is:

\[ \pi_t S^I_t = E_t \int_t^\infty \pi_s \left( \lambda_s a Y_s K_{I,s}^{\beta_I} (L_{I,s}^*)^{1-\beta_I} - w_s L_{I,s} \right) \, ds \]

the intratemporal first order condition with respect to labor implies that:

\[ (1 - \beta_I) \lambda_s a K_{I,s}^{\beta_I} (L_{I,s}^*)^{1-\beta_I} = w_s L_{I,s}^* \]

therefore

\[ \pi_t S^I_t = E_t \int_t^\infty \pi_s \left( \lambda_s a \beta_I Y_s K_{I,s}^{\beta_I} (L_{I,s}^*)^{1-\beta_I} \right) \, ds \]

Moreover, because

\[ \frac{\partial J}{\partial Y_t} Y_t = E_t \int_t^\infty \pi_s \left( \lambda_s a Y_s K_{I,s}^{\beta_I} (L_{I,s}^*)^{1-\beta_I} \right) \, ds \]

this implies

\[ \pi_t S^I_t = \beta_I J_Y Y_t \]

therefore

\[ \frac{S^I_t}{S^C_t} = \frac{\beta_I f''(\omega)}{\beta_C (1 - \gamma) f(\omega) - f'(\omega)}. \]

Also, \( \frac{S^I_t}{S^C_t} \) is increasing in \( \omega \) if

\[ \frac{f'''}{f'} > \frac{f''}{f}. \]

To see that this is the case, note that the above condition implies that \( \frac{f'}{f} \) is strictly decreasing. In the regions where \( f'' < 0 \), the above inequality holds because the LHS is positive while the RHS is negative, so we only need to focus on the case where \( f'' > 0 \), where both sides are negative. Now let's consider the cases where:

1. case \( \frac{f'''}{f'} > \frac{f''}{f} \)
In this case $\frac{f''}{f'}$ is a decreasing function. Meanwhile, the slope of $\frac{f''}{f'}$ depends on whether $\frac{L''}{f'} > \frac{f''}{f'}$ or $\frac{L''}{f'} < \frac{f''}{f'}$. We can exclude the case where

$$\frac{f''}{f'} < \frac{f'}{f}$$

since that would mean that $\frac{f''}{f'}$, which asymptotes to 0 as $\omega \to -\infty$ is increasing with $\omega$. Moreover, the two curves cannot cross, since if $\frac{f''}{f'}$ is below $\frac{f''}{f'}$ that it is decreasing and if $\frac{f''}{f'}$ is above then it is increasing. Thus the only possibility is $\frac{L''}{f'} > \frac{f''}{f'}$.

2. case $\frac{L''}{f'} < \frac{f''}{f'}$

In this case $\frac{L''}{f'}$ is an increasing function. On the other hand, if $\frac{L''}{f'} > \frac{f''}{f'}$ then $\frac{f''}{f'}$ is decreasing. But we can rule out this case because $\lim_{\omega \to -\infty} \frac{L''}{f'} = 0$ and the curves cannot intersect. We can also rule out the other case where $\frac{L''}{f'} > \frac{f''}{f'}$, because this would imply that they are both increasing functions but by $\lim_{\omega \to -\infty} \frac{L''}{f'} = 0$ and $\frac{f''}{f'} \in (1-\gamma, 0)$ this is not possible.

From the above, the only possibility then is that $\frac{L''}{f'} > \frac{f''}{f'}$. 

Proof of Lemma 1. Consider the value of a firm in the C-sector that plans to not invest in the future

$$\pi_{C,t}^V = E_t \int_t^\infty \pi_s \left(X_s(K_{C,t}e^{-\delta(s-t)})^{\beta_C}L_{C,s}^{1-\beta_C} - w_sL_{C,s}\right) ds$$

its labor decision yields

$$(1 - \beta_C)X_s(K_{C,t}e^{-\delta(s-t)})^{\beta_C}L_{C,s}^{1-\beta_C} = w_sL_{C,s}$$

Let $\Lambda_{C,s,t} = \exp(\int_t^s \zeta_{C,u}^d du)$. Now consider a firm who follows the optimal investment policy

$$(1 - \beta_C)X_sK_{C,s}^{\beta_C}L_{C,s}^{1-\beta_C} = w_sL_{C,s}^*$$

dividing through, this yields

$$L_{C,s} = L_{C,s}^* \Lambda_{C,s,t}^{-1}$$
which implies

\[
\pi_t s_{C,t}^V = E_t \int_t^\infty \pi_s \left( \beta_c X_s(K_{C,t} e^{-\delta(s-t)})^\beta_c \Lambda_{C,s,t}^{\beta_c - 1} L_{C,s}^{1-\beta_c} \right) ds
\]

\[
= E_t \int_t^\infty \pi_s \left( \beta_c X_s K_{C,s}^{\beta_c} \Lambda_{C,s,t}^{-1} L_{C,s}^{1-\beta_c} \right) ds
\]

\[
= E_t \int_t^\infty \beta_c(X_s K_{C,s}^{\beta_c} L_{C,s}^{1-\beta_c})^{1-\gamma} \Lambda_{C,s,t}^{-1} ds
\]

\[
= \frac{(X_t K_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} E_t \int_t^\infty e^{-u(s-t)} \beta_c(1-\gamma)L_{C,s}^{1-\beta_c(1-\gamma)} \Lambda_{C,s,t}^{\beta_c(1-\gamma)-1} ds
\]

\[
= \frac{(X_t K_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} g(\omega_t)
\]

The Feynman-Kac theorem implies that \( g(\omega_t) \) can be computed as the solution to the ODE:

\[
\beta_c(1-\gamma)L_{C,s}^{1-\beta_c(1-\gamma)} - u g(\omega) + i^*_c \beta_c(1-\gamma) - 1)g(\omega) + D_\omega g(\omega)
\]

\[\square\]

**Proof of Proposition 3.** First, write out (A.1) as

\[
c(i_{I,t}) K_{I,t} + c(i_{C,t}) K_{C,t} = a Y_t K_{I,t}^{1-\beta_I} L_{I,t}^{1-\beta_I}
\]

\[
c(i_{I,t}) = a Y_t K_{I,t}^{1-\beta_I} L_{I,t}^{1-\beta_I} - c(i_{C,t}) \frac{K_{C,t}}{K_{I,t}}
\]

\[
i_{I,t} = c^{-1}(a e^z L_{I,t}^{1-\beta_I} - c(i_{C,t}) e^\psi)
\]

(A.10)

where

\[
\psi_t \equiv \ln(K_{C,t}) - \ln(K_{I,t})
\]

and

\[
z_t \equiv \ln(Y_t) + (\beta_I - 1) \ln(K_{I,t})
\]

The HJB equation is

\[
0 = \max_{i_t, i_C, L_t, X} \left\{ U(C) + J_{K_I}(i_t - \delta_I) + J_{K_C}(i_C - \delta_C) + \mu X J_X + \frac{1}{2} \sigma_X^2 X^2 J_{XX}
\right.
\]

\[
+ \mu_Y Y J_Y + \frac{1}{2} \sigma_Y^2 Y^2 J_{YY} - \rho J \}
\]

Using the guess

\[
J(X, Y, K_I, K_C) = \frac{(X K_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} f(z, \psi)
\]

130
and writing

\[ L^*_l = l(z, \psi) \]
\[ L^*_c = 1 - l(z, \psi) \]

the HJB equation becomes:

\[ 0 = \min_{i_C, l} \left\{ (1-l)^{1-\beta_C(1-\gamma)} + (\beta_l - 1) f_z(z, \psi) - f_\psi(z, \psi)) c^{-1}(a e^{z L_{1,t}} - c(i_{C,t}) e^{\psi}) \right. \]
\[ + (\delta_l - \delta_C) f_\psi(z, \psi) + (\beta_C(1-\gamma) f(z, \psi) + f_\psi(z, \psi)) i_C \]
\[ \left. + \left( \mu_Y - \frac{1}{2} \sigma_Y^2 + (1-\beta_l) \delta_l \right) f_z(z, \psi) + \frac{1}{2} \sigma_Y^2 \sigma_f^2 f_{zz}(z, \psi) - u f(z, \psi) \right\} \]

where as before

\[ u \equiv \rho + (\gamma - 1)(\mu_X - \beta_C \delta_C) - \frac{1}{2} \gamma (\gamma + 1) \sigma_Y^2 \]

The first order condition with respect to investment imply

\[ ((\beta_l - 1) f_z(z, \psi) - f_\psi(z, \psi)) c^{-1}(a e^{z L_{1,t}} - c(i_{C,t}) e^{\psi}) c'(i_{C,t}) e^{\psi} \]
\[ = (\beta_C(1-\gamma) f(z, \psi) + f_\psi(z, \psi)) \]

If as before we write the relative price of the investment good as \( \lambda \), or in other words write the lagrange multiplier associated with (A.1) as \( \lambda \pi \), the price of the investment good can be recovered (As long as \( c(\cdot) \) satisfies Inada conditions) from either firm’s FOC:

\[ \lambda = \frac{J_{KC}}{U_C c'(i_C^*)}. \]

The first order condition for \( l \) is similar to the baseline case

\[ \lambda Y K_l^\beta l^{-\beta_l} = X K_C^\beta (1-l)^{-\beta_C}. \]

The optimal policy for \( l \) is similar to the baseline case and is a function of \( z \) and \( \psi \) only. It is increasing in \( Y \) as long as \( \lambda(Y)Y \) is increasing in \( Y \), which is generally the case as long as the inverse demand curve for the investment good is sufficiently inelastic. In order for \( \lambda(Y)Y \) to be decreasing in \( Y \) it would have to be the case that a positive shock in the investment sector would actually hurt firm profits in the I-sector.

The state vector \( [z, \psi] \) evolves according to

\[ dz_t = \left( \mu_Y - \frac{1}{2} \sigma_Y^2 + (1-\beta_l) a \delta_l - (1-\beta_l) a e^{z_l^* (z_t, \psi_t)^{1-\beta_l}} \right) dt + \sigma_Y dZ_t^Y \]
\[ d\psi_t = \left( i_{C,t} (e^{\psi_t} + 1) - e^{z_l^* (z_t, \psi_t)^{1-\beta_l}} + \delta_l - \delta_C \right) dt \]
As in the proof of lemma 1, the value of assets in place in the consumption sector equals

\[ S_{C,t}^V = \frac{(X_tK_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} g(z_t, \psi_t) \]

where the function \( g(z_t, \psi_t) \) satisfies the PDE:

\[
\beta_c (1 - I(t, \psi))^{(1-\beta_c)(1-\gamma)} - u g(z, \psi) + i_C^*(z, \psi)(\beta_c(1-\gamma) - 1) g(z, \psi) + D_{z,\psi} h(z, \psi)
\]

The value of assets in place in the investment sector is

\[
S_{I,t}^V = \max_{L_t} \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} \left( aY_t(K_{I,t}e^{-\delta(s-t)})\beta_I \hat{L}_I^{1-\beta_I} - \omega_s \hat{L}_{I,s} \right) ds
\]

\[
= \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} a\beta_I \lambda_s Y_s (K_{I,t}e^{-\delta(s-t)}) \beta_I l(z_s, \psi_s)^{1-\beta_I} e^{(\beta_I-1)\int_t^s \psi_s ds} ds
\]

\[
= \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} a\beta_I \lambda_s K_{I,s} e^{z_s l(z_s, \psi_s)^{1-\beta_I} e^{-\int_t^s \psi_s ds}} ds
\]

\[
= \frac{(X_tK_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} a\beta_I \lambda_s K_{I,s} e^{z_s l(z_s, \psi_s)^{1-\beta_I} e^{-\int_t^s \psi_s ds}} ds
\]

\[
= \frac{(X_tK_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} \mathbb{E}_t \int_t^\infty \frac{\pi_s}{\pi_t} a\beta_I Q_s e^{z_s l(z_s, \psi_s)^{1-\beta_I} e^{-\int_t^s \psi_s ds}} ds
\]

\[
= \frac{(X_tK_{C,t}^{\beta_c})^{1-\gamma}}{1-\gamma} \mathbb{E}_t \frac{1}{\pi_t} h(z_t, \psi_t)
\]

where

\[
Q(z, \psi) \equiv \left[ (\beta_I - 1)f_z(z, \psi) - f_\psi(z, \psi), e^{-\psi}(\beta_c(1-\gamma)f(z, \psi) + f_\psi(z, \psi)) \right].
\]

An application of the Feynman-Kac theorem implies that \( h(z, \psi) \) follows the PDE:

\[
a\beta_I e^z Q(z, \psi) l(z, \psi)^{1-\beta_I} - (u + \beta_c(\gamma - 1) i_C^*(z, \psi) + i_C^*(z, \psi)) h(z, \psi) + D_{z,\psi} h(z, \psi)
\]
A.0.3 Numerical Solution

Markov-Chain Approximation

The solution method closely follows Kushner and Dupuis (1993). After replacing the guess for \( J \), the HJB equation becomes:

\[
0 = \min_i \left\{ (1 - l)^{(1 - \gamma)(1 - \beta_C)} - (u + \beta_C (\gamma - 1) a e^{\omega_1 l^{1 - \beta_1}} f(\omega) + f'(\omega)(\mu_Y + \delta - \frac{1}{2} \sigma_Y^2 - a e^{\omega_1 l^{1 - \beta_1}}) + \frac{1}{2} \sigma_Y^2 f''(\omega) \right\}
\]

and \( \omega \) follows

\[
d\omega = (\mu_Y + \delta - \frac{1}{2} \sigma_Y^2 - a e^{\omega_1 l^{1 - \beta_1}}) dt + \sigma_Y dZ_t
\]

One can discretize the state space, creating a grid for \( \omega \) and \( f \) with \( h = \Delta \omega \). Then the following approximations can be used

\[
f'(\omega_n) \approx \frac{f_{n+1} - f_{n-1}}{2h}
\]

and

\[
f''(\omega_n) \approx \frac{f_{n+1} + f_{n-1} - 2f_n}{h^2}
\]

I then approximate the HJB equation as

\[
f_n = \min_i \left\{ e^{-\beta(\omega_n; l) \Delta t^h} \left[ p_-(\omega_n; l) f_{n-1} + p_+(\omega_n; l) f_{n+1} \right] + (1 - l)^{(1 - \gamma)(1 - \beta_C)} \Delta t^h \right\} \tag{A.11}
\]

where

\[
\beta(\omega_n; l) = u + \beta_C (\gamma - 1) a e^{\omega_1 l^{1 - \beta_1}}
\]

\[
p_-(\omega_n; l) = \frac{1}{2} + h \frac{\mu_Y + \delta - \frac{1}{2} \sigma_Y^2 - a e^{\omega_1 l^{1 - \beta_1}}}{2 \sigma_Y^2}
\]

\[
p_+(\omega_n; l) = \frac{1}{2} - h \frac{\mu_Y + \delta - \frac{1}{2} \sigma_Y^2 - a e^{\omega_1 l^{1 - \beta_1}}}{2 \sigma_Y^2}
\]

\[
\Delta t^h = \frac{h^2}{\sigma_Y^2}
\]

and I have used the approximation

\[
\frac{1}{1 + \beta(\omega_n; l) \Delta t^h} \approx e^{-\beta(\omega_n; l) \Delta t^h}
\]

This corresponds to an MC approximation to \( \omega \), where

\[
p(\omega = \omega_n + h | \omega = \omega_n) = p_+(\omega_n; l)
\]
and

\[ p(\omega = \omega_n - h|\omega = \omega_n) = p_{-}(\omega_n;l) \]

are the transition probabilities and the time interval is \( \Delta t^h \). The Markov chain is locally consistent because

\[
E(\Delta \omega_n|\omega_n) = (\mu_Y + \delta - \frac{1}{2}\sigma^2_Y - ae^\omega_n t^{1-\beta} t) \Delta t^h
\]

\[
E((\Delta \omega_n - E\Delta \omega_n)^2|\omega_n) = \sigma^2_Y \Delta t^h + o(\Delta t^h)
\]

Note that care must be taken when choosing \( h \) to ensure that the probabilities are non-negative for all admissible controls \( l \in [0,1] \) at all points in the grid. Alternative differencing schemes that produce positive probabilities can also be used.

Using an initial guess for \( f \), say \( f_0 \), one can numerically compute the minimum in (A.11). Then, given \( l_n^* \), one can start from \( n = 0 \) and recursively compute the update on \( f \) using the Gauss-Seidel algorithm:

\[
f_{n+1} = e^{-\beta(\omega_n; l_n^*) \Delta t^h} \left[ p_{-}(\omega_n; l_n^*) f_{n-1}^+ + p_{+}(\omega_n; l_n^*) f_{n+1}^+ \right] + (1 - l_n^*) (1-\gamma)(1-\beta c) \Delta t^h \quad (A.12)
\]

I impose a reflecting barrier on \( \omega \) at the boundaries of the grid. This reduces to \( f_0 = f_1 \) and \( f_N = f_{N-1} \), since there is no discounting at the boundary and

\[
p(\omega = \omega_0 + h|\omega = \omega_0) = 1
\]

\[
p(\omega = \omega_N - h|\omega = \omega_N) = 1
\]

\[
\Delta t^h(\omega_N) = \Delta t^h(\omega_0) = 0
\]

Finally, because the minimum in (A.11) is costly to compute, I iterate a couple of times on (A.12) before updating the policy function.
Solution and comparative statics

I pick parameters to approximately match the first two moments of consumption and the relative size of the two sectors. I pick $\beta_C = 0.3$, $\beta_I = 0.3$, $\delta = 0.06$, $\mu_Y = 0.04$, $\sigma_Y = 0.1$, $\mu_X = 0.03$, $\sigma_X = 0.01$, $\gamma = 15$. As the adjustment cost function, I pick

$$c(i) = i + \frac{1}{2} i^2.$$  

When inverting $c(\cdot)$, I pick the larger root. The unconditional moments of returns and real variables produced by the model are shown below.

Table A.1: Unconditional Moments

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>Actual(1951:2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\Delta \log(C))$</td>
<td>3.48%</td>
<td>3.48%</td>
</tr>
<tr>
<td>$\sigma(\Delta \log(C))$</td>
<td>1.68%</td>
<td>1.68%</td>
</tr>
<tr>
<td>$E(R_{MKT} - r_f)$</td>
<td>1.69%</td>
<td>7.06%</td>
</tr>
<tr>
<td>$E(R_{IMC})$</td>
<td>-0.55%</td>
<td>-2.39%</td>
</tr>
<tr>
<td>$E(R_{VMG})$</td>
<td>1.26%</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{MKT}$</td>
<td>21.43%</td>
<td>14.70%</td>
</tr>
<tr>
<td>$\sigma_{IMC}$</td>
<td>9.55%</td>
<td>12.68%</td>
</tr>
<tr>
<td>$\sigma_{VMG}$</td>
<td>20.44%</td>
<td></td>
</tr>
<tr>
<td>$E \frac{MKCAP_t}{HKCAP_t}$</td>
<td>0.09</td>
<td>0.05-0.12</td>
</tr>
</tbody>
</table>
The graphs above plot the numerical solution of the model. The state variable is $\omega \equiv \ln Y - \ln K_C$, where $Y$ is the I-shock and $K_C$ is the capital stock in the consumption sector. The graphs plot the normalized value function, $f(\omega)$, the optimal labor allocation policy, $l(\omega)$, the normalized profitability in the I-sector, $\lambda Y$, Tobin’s $q$, the covariance of consumption with the I-shock, the value of the market portfolio, $S^M$, the relative value of the two sectors, $S^I/S^C$ and the relative value of assets in place vs growth opportunities in the C-sector, $S^V/S^G$. The parameters are picked to match the first two moments of consumption growth and the relative size of the two sectors. I pick $\beta_C = 0.3$, $\beta_Y = 0.3$, $\delta = 0.06$, $\mu_Y = 0.04$, $\sigma_Y = 0.2$, $\mu_X = 0.03$, $\sigma_X = 0.01$, $a = 0.03$, $\gamma = 4$. 

Figure A-1: Solution of the baseline model
The figures above plot the normalized value function, $f(z, \psi)$, the investment policy threshold $\hat{\psi}(z)$, the optimal labor allocation to the I-sector, $l(z, \psi)$, the stationary distribution of the state variables, $p(z, \psi)$, the relative value of the investment versus the consumption sector $S^I / S^C$ and the relative value of assets in place versus future growth opportunities $S^V / S^G$ for the whole economy. The state variables are the marginal productivity of capital in the I-sector, $z \equiv \ln Y - (1 - \beta_I) \ln K_I$, and the ratio of two capital stocks, $\psi \equiv \ln K_C - \ln K_I$. 

137
A.0.4 Data

IMC

I use the 1997 BEA Standard Make and Use Tables at the detailed level. I use the standard make (table 1) and use (table 2) tables. The uses tables enumerates the contribution of each IO commodity code to Personal Consumption Expenditures (IO code F01000) and Gross Private Fixed Investment (IO code F02000). I use the make tables along with the NAICS-IO map to construct a mapping between 6-digit NAICS Codes to IO commodity codes. I then use the uses table to create a map from IO codes to Investment or Consumption. Because some industries contribute to both PCE and GPFI, I follow two schemes to create a unique link. The first assigns industries to the sector they contribute the most value in terms of producer’s prices excluding transportation costs. The second scheme classifies industries as consumption or investment if they contribute only in one sector.

I use COMPUSTAT to create a PERMNO-NAICS link and form value weighted portfolios using simple returns on all common stocks traded on NYSE, AMEX and Nasdaq. I construct two portfolios, using the first (IMC) and the second (IMCX) classification scheme. Examples of Investment industries are

<table>
<thead>
<tr>
<th>IO Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>213111</td>
<td>Drilling oil and gas wells</td>
</tr>
<tr>
<td>333111</td>
<td>Farm machinery and equipment manufacturing</td>
</tr>
<tr>
<td>333295</td>
<td>Semiconductor machinery manufacturing</td>
</tr>
<tr>
<td>334111</td>
<td>Electronic computer manufacturing</td>
</tr>
<tr>
<td>334220</td>
<td>Broadcast and wireless communications equipment</td>
</tr>
<tr>
<td>336120</td>
<td>Heavy duty truck manufacturing</td>
</tr>
</tbody>
</table>

Examples of Consumption industries are

<table>
<thead>
<tr>
<th>IO Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1111B0</td>
<td>Grain farming</td>
</tr>
<tr>
<td>221100</td>
<td>Power generation and supply</td>
</tr>
<tr>
<td>311410</td>
<td>Frozen food manufacturing</td>
</tr>
<tr>
<td>312110</td>
<td>Soft drink and ice manufacturing</td>
</tr>
<tr>
<td>325611</td>
<td>Soap and other detergent manufacturing</td>
</tr>
<tr>
<td>334300</td>
<td>Audio and video equipment manufacturing</td>
</tr>
</tbody>
</table>

The full list of IO codes and their assignments into industries is available from the author’s website.

10 IMC sorted portfolios

The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq, in the CRSP Monthly Stock Database. For every stock, I estimate pre-ranking covariances
on IMC based on a 5-year window using weekly log excess returns. I sort stocks into IMC co-
variance deciles. I construct value-weighted portfolios using simple monthly returns, tracking
the portfolios for 5 years until next re-balancing.

24 IND/IMC sorted portfolios

The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq,
in the CRSP Monthly Stock Database. For every stock, I estimate pre-ranking covariances
on IMC based on a 5-year window using weekly log excess returns. Stocks are sorted into
eight industries based on their two-digit SIC codes: (1) nondurables manufacturing, (2)
durables manufacturing, (3) other manufacturing, (4) nondurables retail, (5) durables retail,
(6) services, (7) finance, and (8) natural resource. Within each industry, stocks are then
sorted into three portfolios based on their IMC covariance using breakpoints of 30th and 70th
percentiles. I construct value-weighted portfolios using simple monthly returns, tracking the
portfolios for 5 years until next re-balancing.

25 MKT/IMC sorted portfolios

The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq,
in the CRSP Monthly Stock Database. For every stock, I estimate pre-ranking betas on
excess returns on the CRSP Value-Weighted Index and IMC based on a 5-year window using
weekly log excess returns. Using December market capitalization from CRSP and the pre-
ranking covariances, including only stocks that have full observations, I first sort stocks into
BM quintiles using NYSE breakpoints and then into IMC covariance quintiles. I construct
value-weighted portfolios using simple monthly returns, tracking the portfolios for 5 years
until next re-balancing.

25 ME/IMC sorted portfolios

The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq,
in the CRSP Monthly Stock Database. For every stock, I estimate pre-ranking covariances
on IMC based on a 5-year window using weekly log excess returns. Using December market
capitalization from CRSP and the pre-ranking covariances, including only stocks that have
full observations, I first sort stocks into BM quintiles using NYSE breakpoints and then
into IMC covariance quintiles. I construct value-weighted portfolios using simple monthly
returns, tracking the portfolios for 5 years until next re-balancing.

25 BM/IMC sorted portfolios

The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq,
in the CRSP Monthly Stock Database. For every stock, I estimate pre-ranking covariances
on IMC based on a 5-year window using weekly log excess returns. Using book to market
equity from COMPUSTAT and the pre-ranking covariances, including only stocks that have
full observations, I first sort stocks into BM quintiles using NYSE breakpoints and then into
IMC covariance quintiles. Portfolios are formed in June every 5 years. I construct value-weighted portfolios using simple monthly returns, tracking the portfolios for 5 years until next re-balancing.

24 BM/Industry portfolios

The construction of these portfolios follows Yogo (2006). The portfolios are formed from ordinary common equity, traded in NYSE, AMEX, or Nasdaq, in the CRSP Monthly Stock Database. In June of each year $t$, stocks are sorted into eight industries based on their two-digit SIC codes: (1) nondurables manufacturing, (2) durables manufacturing, (3) other manufacturing, (4) nondurables retail, (5) durables retail, (6) services, (7) finance, and (8) natural resource. Within each industry, stocks are then sorted into three levels of book-to-market equity using breakpoints of 30th and 70th percentiles, based on their value in December of $t-1$. The book equity data is a merge of historical data from Moodys Manuals (available from Kenneth Frenchs web page) and COMPUSTAT. See Davis, Fama, and French (2000) for details on the computation of book equity.

25 Size/BM and 30 Industry portfolios

The data come from Kenneth French’s web page.

Macroeconomic Data

Monthly data come from the website of the St Louis Fed. Quarterly quantity data and price deflators come from the website of the Bureau of Economic Analysis, specifically from the NIPA tables 5.3.3, 5.3.4, 2.3.3, 1.1.3 and 1.1.4.

A.0.5 Empirical Methodology

GMM Cross-Sectional tests

I estimate the linear factor model using two-step GMM. As Cochrane (2001) illustrates, if one is interested in estimating a linear model for the SDF of the following form:

$$m = a - bF$$

One can test the moment restriction

$$E[mR^e] = 0. \quad (A.13)$$

Letting the vector of unknown parameters $\theta = [b, \mu_F]$ and the data $X_t = [R^e_t, F_t]$, where $F$ is the factors and $R^e$ are excess returns. If one is using excess returns, then the mean of the pricing kernel is unidentified. If the model is correct, then (A.13) must hold at the true parameter values:

$$E[g(X, \theta_0)] = 0$$
where

\[ g(X, \theta) = \begin{bmatrix} R_t^e - R_t^e (F_t - \mu_F)'b \\ F_t - \mu_F \end{bmatrix} \]

I use the first stage weighting matrix

\[ W = \begin{bmatrix} kI_N & 0 \\ 0 & \hat{\Sigma}_{ff}^{-1} \end{bmatrix} \]

Following Gomes, Kogan and Yogo (2006) I pick \( k = \text{det}(\hat{\Sigma}_R)^{-1/N} \). The first stage weighting matrix puts equal weight in each of the N asset pricing restrictions. I compute the spectral density matrix for the second stage using the Newey-West estimator with 3 lags of returns.

I use the covariance rather than the beta representation because I am interested in the marginal ability of IMC to price each cross-section. As Cochrane 2001 illustrates, one can test whether a factor is priced, given the other factors in the specification by \( b \neq 0 \).

**Fama-McBeth Cross-Sectional tests**

I run Fama and MacBeth (1973) cross-sectional regressions of simple monthly returns of all NYSE, AMEX, and Nasdaq stocks on covariances and characteristics. Covariances are estimated using a procedure similar to Fama and French (1992). Specifically, for each individual stock I estimate the covariance of its returns with IMC using 5 years of weekly log excess returns. At the end of a five year period, stocks are then sorted into 100 pre-ranking covariance centiles. I then compute the equal-weighted monthly log excess returns on these 100 portfolios over the next 5 years. This procedure is repeated every 5 years, forming a time-series of returns on these 100 portfolios. I then re-estimate covariances for the portfolios formed from the pre-ranking sorts using 5 years of monthly data to obtain post-ranking covariances. The post-ranking covariance estimate for a given portfolio is then assigned to each stock in the portfolio. Portfolio assignments are updated every 5 years. Every month the cross-section of stock returns in excess of risk free rate is then regressed on a constant, the covariance with the excess return on the CRSP value-weighted index, the covariance with the return on the IMC portfolio, the log of market capitalization (ME) on December of year t-1 and the log of Book to Market (BM) of year t-1.
Appendix B

Chapter 3: Appendix

B.0.6 Proofs

This section contains the unproven propositions from sections 3 to 5 of the paper.

Proof of Proposition 2: We proceed by backward induction. If $R = 0$, it follows by assumption that $M^* = M_L$ and $X_B^* = 0$ and from Proposition 1 that $X_A^* = \min(X_A^{FB}, M_L)$. Therefore,

$$X_A^* = \begin{cases} \frac{1}{4} \Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ M_L & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases},$$

and

$$\Delta p^* = \begin{cases} \frac{1}{2} \Delta X_N & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \lambda \Delta X_N - 2\lambda M_L & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases}.$$

Payoffs are given by:

$$\Pi_A^* = \begin{cases} \frac{3}{8} \Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\ \lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L \end{cases}.$$

Now assume that $R = 1$. Given $\Delta X_N$ and $X_B$, the arbitrageur chooses his demand such that:

$$\max_{X_A \leq M_L} \lambda X_A (\Delta X_N - 2X_A - 2X_B).$$

This implies that:

$$X_A^* = \begin{cases} \frac{1}{4} \Delta X_N - \frac{1}{2}X_B & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2} \Delta X_N - 2M_L \\ M_L & \text{if } R = 1 \text{ and } X_B \leq \frac{1}{2} \Delta X_N - 2M_L \end{cases}.$$
As a result, given $\Delta X_N$, the bank's payoff as a function of its own choice of trade is:

$$\Pi_B(X_B) = \begin{cases} 
-2\lambda (X_B - \frac{1}{4} \Delta X_N + \frac{1}{2} M_L) 2+ \frac{3}{8} \Delta X_N 2 - \frac{3}{2} M_L \Delta X_N + \frac{3}{2} M_L 2 & \text{if } R = 1 \text{ and } X_B \leq \frac{1}{2} \Delta X_N - 2 M_L \\
-\lambda (X_B - \frac{1}{4} \Delta X_N) 2 + \frac{1}{16} \Delta X_N 2 & \text{if } R = 1 \text{ and } X_B \geq \frac{1}{2} \Delta X_N - 2 M_L
\end{cases}$$

This piecewise quadratic objective consists of two parts. For small values of $X_B$, the arbitrageur is constrained while, for higher values of $X_B$, he is unconstrained. The quadratic function corresponding to the constrained segment reaches its maximum at $X_B = \frac{1}{4} \Delta X_N - \frac{1}{2} M_L$ while the unconstrained version achieves its maximum at a strictly higher point, namely $X_B = \frac{1}{4} \Delta X_N$. We now determine the bank’s optimal $X^*_B$ given $R = 1$ and a particular $\Delta X_N$. To do so, we consider 3 regions of $\Delta X_N$:

**Region 1:** $\Delta X_N \leq 6 M_L$. We have $\frac{1}{4} \Delta X_N - \frac{1}{2} M_L \geq \frac{1}{2} \Delta X_N - 2 M_L$ which implies that there is only one local/global maximum in the bank’s objective function, namely $X^*_B = \frac{1}{4} \Delta X_N$. Notice that this implies the arbitrageur is unconstrained (i.e. $X_A \leq M_L$).

**Region 2:** $\Delta X_N \geq 8 M_L$. We have $\frac{1}{4} \Delta X_N \leq \frac{1}{2} \Delta X_N - 2 M_L$ and, therefore, the only local/global maximum in the bank’s objective is $X^*_B = \frac{1}{4} \Delta X_N - \frac{1}{2} M_L$. The arbitrageur is constrained.

**Region 3:** $6 M_L \leq \Delta X_N \leq 8 M_L$. In this region, there are two local maxima in the bank’s objective: $X_B = \frac{1}{4} \Delta X_N$ (constrained) and $X_B = \frac{1}{4} \Delta X_N - \frac{1}{2} M_L$ (unconstrained). The global maximum is given by the constrained case if and only if

$$\frac{1}{8} \Delta X_N 2 - \frac{1}{2} M_L \Delta X_N + \frac{1}{2} M_L 2 \geq \frac{1}{16} \Delta X_N 2,$$

or equivalently,

$$\Delta X_N \geq (4 + 2\sqrt{2}) M_L.$$

Therefore, we have:

$$X^*_B = \begin{cases} 
\frac{1}{4} \Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2}) M_L \\
\frac{1}{4} \Delta X_N - \frac{1}{2} M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2}) M_L
\end{cases}$$

and

$$X^*_A = \begin{cases} 
\frac{1}{8} \Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2}) M_L \\
M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2}) M_L
\end{cases}$$

Meanwhile, price spreads and arbitrageur profits are given by:

$$\Delta p^* = \begin{cases} 
\frac{1}{4} \Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2}) M_L \\
\frac{1}{2} \Delta X_N - \lambda M_L & \text{if } R = 1 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2}) M_L
\end{cases}$$
and
\[
\Pi_A^* = \begin{cases} 
\frac{1}{32} \Delta X_N & \text{if } R = 1 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\
\frac{1}{2} M_L \Delta X_N - \lambda M_L^2 & \text{if } R = 1 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2})M_L 
\end{cases}
\]

It is easy to verify that \( \Pi_A^*(R = 0) \geq \Pi_A^*(R = 1) \) and, hence, \( R^* = 0 \). Q.E.D.

**Corollary 3:** If the arbitrageur doesn’t reveal his information to the bank, price spreads are given by:

\[
\Delta p^* = \begin{cases} 
\frac{1}{2} \Delta X_N & \text{if } \Delta X_N \leq 4M_L \\
\lambda \Delta X_N - 2\lambda M_L & \text{if } \Delta X_N \geq 4M_L 
\end{cases}
\]

Further, the arbitrageur and bank payoffs are:

\[
\Pi_A^* = \begin{cases} 
\frac{1}{8} \Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq 4M_L \\
\lambda M_L \Delta X_N - 2\lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq 4M_L 
\end{cases}
\quad \text{and} \quad \Pi_B^* = 0.
\]

**Remark for Corollary 3.** So long as the arbitrageur is unconstrained, the outcome is as in the first-best equilibrium. Otherwise, he binds to his default financial constraint and has a fixed convergence effect on the price spread (reducing it by \( 2\lambda M_L \)).

**Corollary 4:** If the arbitrageur does reveal his information to the bank, price spreads are given by:

\[
\Delta p^* = \begin{cases} 
\frac{1}{4} \Delta X_N & \text{if } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\
\frac{1}{2} \Delta X_N - \lambda M_L & \text{if } \Delta X_N \geq (4 + 2\sqrt{2})M_L 
\end{cases}
\]

Further, the arbitrageur and bank payoffs are:

\[
\Pi_A^* = \begin{cases} 
\frac{1}{8} \Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\
\frac{1}{2} M_L \Delta X_N - \lambda M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2})M_L 
\end{cases}
\quad \text{and} \quad \Pi_B^* = \begin{cases} 
\frac{1}{16} \Delta X_N^2 & \text{if } R = 0 \text{ and } \Delta X_N \leq (4 + 2\sqrt{2})M_L \\
\frac{1}{8} \Delta X_N^2 - \frac{1}{2} M_L \Delta X_N + \frac{1}{2} M_L^2 & \text{if } R = 0 \text{ and } \Delta X_N \geq (4 + 2\sqrt{2})M_L 
\end{cases}
\]

respectively.

**Remark for Corollary 4.** When the initial mispricing is small, the arbitrageur is unconstrained following the bank’s aggressive hold-up and the price spread is half that of the first-best equilibrium. Meanwhile, if \( \Delta X_{N,t} \) is large enough, the arbitrageur is still constrained and the price spread
equals half that of the one-shot game. In both cases though, equilibrium mispricing is smaller than in the first-best. This is due to the fact that the bank is unconstrained and doesn’t internalize the arbitrageur’s welfare (i.e. a standard competition effect).

**Proof of Lemma 5:** First of all, notice that if $X_{B,t} > 0$, then an unconstrained arbitrageur would choose $X_{A,t} > \frac{1}{4} \Delta X_{N,t} - X_{B,t}$ since his unwillingness to internalize the effect of his demands on the bank’s profits would lead to total demands that exceed the first-best level. This implies that if the arbitrageur is constrained and $X_{B,t} + M_t = \frac{1}{4} \Delta X_{N,t}$, then $X_{A,t} = M_t$.

We now proceed to prove this lemma by contradiction. Assume there exists an efficient equilibrium such that $R^*_t = 1$ and $X^*_{B,t} + M^*_t \neq \frac{1}{4} \Delta X_{N,t}$ for some $\Delta X_{N,t}$. It follows that $\Pi^*_A,t + \Pi^*_B,t < \frac{1}{8} \Delta X_{N,t}^2$. Consider an alternative set of strategies identical to this one with the exception that, at the aforementioned $\Delta X_{N,t}$, we have $X^*_{B,t} = \frac{2\Pi^*_B,t}{\lambda \Delta X_{N,t}}$ and $M^*_t = \frac{1}{4} \Delta X_{N,t} - X^*_{B,t}$. This implies that $\Pi^*_A,t + \Pi^*_B,t = \frac{1}{8} \Delta X_{N,t} > \Pi^*_A,t + \Pi^*_B,t$ or equivalently, since $\Pi^*_B,t = \Pi^*_B,t$, that $\Pi^*_A,t > \Pi^*_A,t$. Notice that since the initial equilibrium satisfied both the arbitrageur’s IR constraint and the bank’s IC constraint, so does the new one (the IC constraint still holds since the value of the relationship $V^*_t$ in the new construction is equal to the $V^*_t$ from the initial equilibrium). This contradicts the claim that the SPE is efficient. Q.E.D.

**Proof of Proposition 6:** We consider three different regions for relationship values: (i) $V^*_t \leq \lambda M_L / 2$, (ii) $\lambda M_L / 2 < V^*_t \leq (1 + 2\sqrt{2}) \lambda M_L / 2$, and (iii) $V^*_t > (1 + 2\sqrt{2}) \lambda M_L / 2$.

**First Region:** If $V^*_t \leq \lambda M_L / 2$, it is impossible to satisfy both the arbitrageur’s IR and the bank’s IC constraints at $\Delta X_{N,t} = 4M_L$. As a result, no communication can ever be sustained between the two and $\delta x^*_t = 0$.

**Second Region:** In this case, the breakdown of communication happens before $\Delta X_{N,t} = (4 + 2\sqrt{2}) M_L$ and $\delta x^*_t$ solves:

$$
\frac{1}{4} \delta x^*_t - 2M_L + 4M_L^2 \left( \frac{1}{\delta x^*_t} \right) = \frac{1}{8} \delta x^*_t - \frac{2V^*}{\lambda} \left( \frac{1}{\delta x^*_t} \right).
$$

(B.1)

**Third Region:** When $V^*_t \geq (1 + 2\sqrt{2}) \lambda M_L / 2$, the breakdown occurs after $\Delta X_{N,t} = (4 + 2\sqrt{2}) M_L$ and $\delta x^*_t$ solves:

$$
\frac{1}{4} \delta x^*_t - 2M_L + 4M_L^2 \left( \frac{1}{\delta x^*_t} \right) = \frac{1}{4} \delta x^*_t - M_L + \left( M_L - \frac{2V^*}{\lambda} \right) \left( \frac{1}{\delta x^*_t} \right).
$$

(B.2)

Q.E.D.

**Proof of Lemma 7:** We split this proof into two components: (i) show that $R^*_t = 1$ for $\Delta X_{N,t} = \delta x^*_t$, and (ii) show that if $R^*_t = 1$ for $\Delta X_{N,t}$ and $4M_L \leq \Delta X_{N,t} \leq \Delta X_{N,t}$, then $R^*_t = 1$ for $\Delta X_{N,t}$. Both are proven by contradiction.

**First Component:** Suppose there exists a Pareto-optimal equilibrium where $R^*_t = 0$ for $\Delta X_{N,t} = \delta x^*_t$. Consider a strategy profile that is identical to this equilibrium with the exception that it
differs at the point $\Delta X_{N,t} = \delta x_t^*$. At that point, set $X_{B,t}' = \frac{2(M_t - \Pi_{A,t})}{\lambda \delta x_t^*}$ and $M_t' = \frac{2\Pi_{A,t}}{\lambda \delta x_t^*}$. It is easy to check that both the arbitrageur’s IR and the bank’s IC constraints are met for $\Delta X_{N,t} = \delta x_t^*$ and for all $\Delta X_{N,t}$ where $R^*_t = 1$. Therefore, the new strategy is a subgame perfect equilibrium that Pareto-dominates the aforementioned one. Contradiction.

Second Component: Assume that there exists a Pareto-optimal equilibrium where $R^*_t = 1$ for $\Delta X_{N,t}$ and $R^*_t = 0$ for $\Delta X_{N,t}'$. Since $R^*_t = 1$, we have $X_t > X_t'$ for $\Delta X_{N,t}$. It follows that $X_t > X_t'$ for $\Delta X_{N,t}'$ as well. Pick any $X_B \in [X_t', X_t]$ and consider a strategy profile that is identical to the stated equilibrium except that it sets $R_t' = 1$ and $X_{B,t} = X_B$ (while satisfying Lemma 5) for $\Delta X_{N,t}'$. It is easily verified that both the arbitrageur’s IR and the bank’s IC constraints are satisfied for $\Delta X_{N,t}$ and for all $\Delta X_{N,t}$ where $R^*_t = 1$. This new strategy profile is subgame perfect and Pareto-dominates the aforementioned equilibrium. Contradiction. Q.E.D.

Proof of Proposition 10: We proceed to prove this by contradiction. Assume there exists two equilibria with $M_L^t < M_L''$ such that $\delta_t^*(M_L^t) \leq \delta_t^*(M_L'')$. From Proposition 6, $\delta_t^*(M_L^t) \leq \delta_t^*(M_L'')$ implies that $V_t^*(M_L^t) < V_t^*(M_L'')$. Therefore, the bargaining power assumption implies that $\Pi_t^*(M_L^t) > \Pi_t^*(M_L'')$ in the interval $4M_L \leq \Delta X_{N,t} \leq \delta x_t^*(M_L^t)$. Now consider an alternative relational contract for $M_L^t$ with:

$$\Pi_t^{**}(M_L^t) = \begin{cases} \Pi_t^*(M_L^t) & \text{if } \Delta X_{N,t} \leq \delta_t^*(M_L^t) \\ \Pi_t^*(M_L'') & \text{if } \Delta X_{N,t} > \delta_t^*(M_L^t) \end{cases}$$  \hspace{1cm} \text{(B.3)}

It follows that $V_t^{**}(M_L^t) > V_t^*(M_L'')$ and the bank’s IC constraint is strictly satisfied up until $\delta_t^*(M_L'')$. Given optimal collusion, the arbitrageur’s IR constraint is also met up till $\delta_t^*(M_L'')$ since his outside option is strictly lower when $M_L = M_L^t$ than when $M_L = M_L''$. Continuity of $\Pi_t^*$ and the strict satisfaction of the bank’s IC under $\Pi_t^{**}(M_L^t)$ implies that there exists an $\epsilon > 0$ such that we can construct a relational contract with:

$$\Pi_t^{***}(M_L^t) = \begin{cases} \Pi_t^*(M_L^t) & \text{if } \Delta X_{N,t} \leq \delta_t^*(M_L^t) \\ \Pi_t^*(M_L'') & \text{if } \delta_t^*(M_L^t) < \Delta X_{N,t} \leq \delta_t^*(M_L'') \\ \max_{\Delta X_{N,t}} \Pi_t^*(M_L'') & \text{if } \delta_t^*(M_L'') < \Delta X_{N,t} \leq \delta_t^*(M_L'') + \epsilon \\ 0 & \text{if } \Delta X_{N,t} > \delta_t^*(M_L'') + \epsilon \end{cases}$$  \hspace{1cm} \text{(B.4)}

This contract satisfies $\delta x_t^{***}(M_L^t) > \delta x_t^*(M_L'')$ and contradicts the statement that the equilibrium with $\delta x_t^*(M_L^t)$ is efficient. Q.E.D.
Lemma 11: A monopolist bank’s cooperation profit, as a function of $V_t^*$, is given by:

$$\Pi_{B,t}^{c,M} = \begin{cases} 
0 & \text{if } \Delta X_{N,t} \leq 4M_L \\
\frac{1}{8} \Delta X_{N,t}^2 - \lambda M_L \Delta X_{N,t} + 2\lambda M_L^2 & \text{if } 4M_L \leq \Delta X_{N,t} \leq \delta x_t^* \\
0 & \text{if } \Delta X_{N,t} \geq \delta x_t^*
\end{cases}$$

where $\delta x_t^*$ depends on $V_t^*$ as described in Proposition 6.

Proof of Lemma 11: When $V_t^* \leq \Pi_{B,t}^d$ and $\Delta X_{N,t} \leq \delta x_t^*$, the bank cannot commit to give the arbitrageur all the surplus, but can commit to cooperation. In this case, $\Pi_{B,t}^c = \Pi_{B,t}^d - V_t^*$. We apply this logic to different regions that span all possible values of $V_t^*$.

Region 1: $V_t^* \leq \lambda M_L^2$. There is no communication between the two parties and $\Pi_{B,t}^{c,PC}$ is always zero.

Region 2: $\lambda M_L^2 \leq V_t^* \leq (\frac{1+2\sqrt{2}}{2}) \lambda M_L^2$. In this region, the bank cannot commit not to trade starting at $\Delta X_{N,t} = 4\sqrt{V_t^*/\lambda} \geq 4M_L$. Meanwhile, since $\delta x_t^* \leq (4 + 2\sqrt{2})M_L$ so the bank’s profit given that $\Delta X_{N,t}$ is between $4\sqrt{V_t^*/\lambda}$ and $\delta x_t^*$ is $\frac{1}{16} \Delta X_{N,t}^2 - \frac{1}{2} M_L \Delta X_{N,t} + \frac{3}{2} M_L^2 - V_t^*$. Otherwise it is 0.

Region 3: $\left(\frac{1+2\sqrt{2}}{2}\right) \lambda M_L^2 \leq V_t^* \leq \left(\frac{3+2\sqrt{2}}{2}\right) \lambda M_L^2$. This is similar to region 2, with the exception that $\delta x_t^* \geq (4 + 2\sqrt{2})M_L$, so the bank’s profit given that $\Delta X_{N,t}$ is between $(4 + 2\sqrt{2})M_L$ and $\delta x_t^*$ is $\frac{1}{8} \Delta X_{N,t}^2 - \frac{1}{2} M_L \Delta X_{N,t} + \frac{3}{2} M_L^2 - V_t^*$.

Region 4: $V_t^* \geq (\frac{3+2\sqrt{2}}{2}) \lambda M_L^2$. Similar to region 3 but the point where the bank starts trading is now above $(4 + 2\sqrt{2})M_L$ and equals $2M_L + 2\sqrt{2M_L^2 + \frac{3V_t^*}{\lambda}}$. Therefore, whenever the bank makes positive profits, it makes $\frac{1}{8} \Delta X_{N,t}^2 - \frac{1}{2} M_L \Delta X_{N,t} + \frac{3}{2} M_L^2 - V_t^*$. Q.E.D.

Lemma 12: The bank’s cooperation profits, as a function of $V_t^*$, are given by:

Region 1: $V_t^* \leq \lambda M_L^2$.

$$\Pi_{B,t}^{c,PC} = 0$$

Region 2: $\lambda M_L^2 \leq V_t^* \leq \left(\frac{1+2\sqrt{2}}{2}\right) \lambda M_L^2$.

$$\Pi_{B,t}^{c,PC} = \begin{cases} 
0 & \text{if } \Delta X_{N,t} \leq 4\sqrt{\frac{V_t^*}{\lambda}} \\
\frac{1}{8} \Delta X_{N,t}^2 - V_t^* & \text{if } 4\sqrt{\frac{V_t^*}{\lambda}} \leq \Delta X_{N,t} \leq 8M_L - 4\sqrt{2M_L^2 - \frac{V_t^*}{\lambda}} \\
0 & \text{if } \Delta X_{N,t} \geq 8M_L - 4\sqrt{2M_L^2 - \frac{V_t^*}{\lambda}}
\end{cases}$$
Region 3: 
\begin{equation*}
\left( \frac{1 + 2 \sqrt{2}}{2} \right) \lambda M_L 2 \leq V_t^* \leq \left( \frac{3 + 2 \sqrt{2}}{2} \right) \lambda M_L 2.
\end{equation*}

\begin{align*}
\Pi_{B,t}^{PC} &= \begin{cases}
0 & \text{if } \Delta X_{N,t} \leq 4 \sqrt{\frac{V_t^*}{\lambda}} \\
\frac{1}{16} \Delta X_{N,t}^2 - V_t^* & \text{if } 4 \sqrt{\frac{V_t^*}{\lambda}} \leq \Delta X_{N,t} \leq (4 + 2 \sqrt{2}) M_L \\
\frac{1}{8} \Delta X_{N,t}^2 - \frac{3}{2} M_L \Delta X_{N,t} + \frac{3}{2} M_L 2 - V_t^* & \text{if } (4 + 2 \sqrt{2}) M_L \leq \Delta X_{N,t} \leq 3 M_L + \frac{2 V_t^*}{\lambda} \left( \frac{1}{M_L} \right) \\
0 & \text{if } \Delta X_{N,t} \geq 3 M_L + \frac{2 V_t^*}{\lambda} \left( \frac{1}{M_L} \right)
\end{cases}
\end{align*}

Region 4: 
\begin{equation*}
V_t^* \geq \left( \frac{3 + 2 \sqrt{2}}{2} \right) \lambda M_L 2.
\end{equation*}

\begin{align*}
\Pi_{B,t}^{PC} &= \begin{cases}
0 & \text{if } \Delta X_{N,t} \leq 2 M_L + 2 \sqrt{2 M_L 2 + \frac{2 V_t^*}{\lambda}} \\
\frac{1}{8} \Delta X_{N,t}^2 - \frac{3}{2} M_L \Delta X_{N,t} + \frac{3}{2} M_L 2 - V_t^* & \text{if } 2 M_L + 2 \sqrt{2 M_L 2 + \frac{2 V_t^*}{\lambda}} \leq \Delta X_{N,t} \leq 3 M_L + \frac{2 V_t^*}{\lambda} \left( \frac{1}{M_L} \right) \\
0 & \text{if } \Delta X_{N,t} \geq 3 M_L + \frac{2 V_t^*}{\lambda} \left( \frac{1}{M_L} \right)
\end{cases}
\end{align*}

Proof of Lemma 12: The proof follows immediately from Corollary 4 given the fact that: (i) no relationship is sustained if \( V_t^* \leq \lambda M_L 2 \), (ii) the bank starts front-running and the relationship breaks down prior to \( \Delta X_{N,t} = (4 + 2 \sqrt{2}) M_L \) if \( \lambda M_L 2 \leq V_t^* \leq (1 + 2 \sqrt{2}) \lambda M_L 2/2 \), (iii) the bank starts front-running prior to \( (4 + 2 \sqrt{2}) M_L \) and the relationship breaks down after \( (4 + 2 \sqrt{2}) M_L \) when \( (1 + 2 \sqrt{2}) \lambda M_L 2/2 \leq V_t^* \leq (3 + 2 \sqrt{2}) \lambda M_L 2/2 \), and (iv) the bank starts front-running and the relationship breaks down after \( (4 + 2 \sqrt{2}) M_L \) when \( V_t^* \geq (3 + 2 \sqrt{2}) \lambda M_L 2/2 \). Q.E.D.

Proof of Proposition 13: We proceed in a similar fashion to the proof of Proposition 10. Assume that \( V_M^* < V_{PC}^* \) and that profits in both cases are given by \( \Pi_{A,M}^*, \Pi_{B,M}^* \) and \( \Pi_{A,PC}^*, \Pi_{B,PC}^* \). Consider the relational contract with:

\begin{equation*}
\Pi_{B,M}^{**} = \begin{cases}
\Pi_{B,M}^* & \text{if } \Delta X_{N,t} \leq \delta x_M^* \\
\Pi_{B,PC}^* & \text{if } \Delta X_{N,t} > \delta x_{PC}^*
\end{cases}
\end{equation*}

As in Proposition 10, this profit function assures that both the arbitrageur’s IR and the bank’s IC constraints are met (the latter strictly). Therefore, there exists an \( \epsilon > 0 \) such that the relational contract with bank profits given by:
that is self-enforcing and satisfies the surplus allocation rule in the monopoly setting. Since the breakdown in communication in this equilibrium is \( \delta x^*_M + \varepsilon \), we have contradicted the statement that the initial monopoly equilibrium was efficient. Q.E.D.

**Proof of Lemma 14:** We explicitly prove the claim for the monopoly case where \( V_M^* > (1 + 2\sqrt{2})\lambda M_L^2/2 \). The relationship value in this case is determined by the fixed point problem:

\[
V = \frac{\delta}{1 - \delta} \int_{\frac{\beta}{4}}^{\frac{\beta}{4} + \frac{4V}{\sigma}} \left[ \frac{\lambda}{8} x^2 - \frac{\lambda \beta \sigma}{4} + \frac{\lambda \beta^2 \sigma^2}{8} \right] \left( \frac{1}{\sigma} \right) \sqrt{\frac{2}{\pi}} \exp \left[ \frac{1}{2} \left( \frac{x}{\sigma} \right)^2 \right] dx.
\] (B.7)

We can divide by \( \lambda \sigma^2 \) on both sides of this equation and rewrite it as:

\[
\hat{V} = \frac{\delta}{1 - \delta} \int_{\frac{\beta}{4}}^{\frac{\beta}{4} + \frac{4\hat{V}}{\sigma}} \left[ \frac{1}{8} \hat{x}^2 - \frac{\beta}{4} \hat{x} + \frac{\beta^2}{8} \right] \sqrt{\frac{2}{\pi}} \exp \left[ \frac{1}{2} \left( \frac{\hat{x}}{\sigma} \right)^2 \right] d\hat{x}.
\] (B.8)

where \( \hat{x} = x/\sigma \) and \( \hat{V} = V/\lambda \sigma^2 \). It immediately follows that \( \langle \delta x^* \rangle = 3\beta/4 + 8\hat{V}/\beta \) is a constant since \( \hat{V} \) is the value of financial relationships when \( \lambda = \sigma = 1 \). The same procedure can be used to prove that \( \langle \delta \hat{x}^* \rangle \) is constant in the other cases since the bank’s profit function is always quadratic in \( \Delta X_{N,t} \). Q.E.D.

**Remark on Lemma 14:** The irrelevance of \( \lambda \) does not depend on the distribution of \( \Delta X_{N,t} \). Likewise, the statement regarding \( \sigma \) will hold for any distribution whose pdf can be written in the form \( f(x; \sigma) = g(x/\sigma)/\sigma \). Many additional distributions have this property (e.g. exponential distribution).

**Proof of Proposition 15:** Assume the bank decides to set her employment offer to attract an arbitrageur of type \( \Delta X_{N,t} = x \). She will set his (targeted) wage offer at \( W = \Pi_A(x) \) since she has monopoly power and minimizing \( W \) will also maximize her ability to screen out speculators. Translating into an expression of wage as a function of profit, by using \( \Pi = \Pi_M = \lambda x/6 \), yields \( W(\Pi) = 2M_L\sqrt{2\lambda \Pi} - 2\lambda M_L^2 \). The constraint that the wage function be non-decreasing implies that \( W(\Pi) = \sup_{\Pi < \Pi} W(\Pi) \) for all \( \Pi \notin \Omega \).

We now proceed to prove the remainder of the proposition in four parts: (i) We show that arbitrators with opportunity \( \Delta X_{N,t} \leq 4M_L \) are not hired by the bank, (ii) that \( \overline{M} \) is bounded, (iii) that \( \overline{w} = \max_{x \leq \overline{M}} E^G[\Pi(X \cdot \overline{\hat{\Pi}})] \), and (iv) that the wage offer is capped at \( \overline{W} = 4\lambda M_L\overline{M} - 2\lambda M_L^2 \).

**First Part:** The bank chooses to screen out arbitrators of type \( \Delta X_{N,t} \leq 4M_L \) because they are unconstrained and their outside option satisfies \( \Pi_A = \Pi_M \). As a result, hiring these types
would not provide additional profits to the bank and would also act to tighten the screening condition for the speculators (since they would benefit from the wage component that attracts the these arbitrators. Therefore, \( W(\Pi) = 0 \) for \( \Pi \leq 2\lambda M_L^2 \).

**Second Part:** Assume that \( \overline{M} \) is unbounded. If \( W(.) \) is unbounded as well, then it is impossible to screen speculators. Meanwhile, if \( W(.) \) is bounded, arbitrators with opportunities \( \Delta X_{N,t} \) greater than a threshold, \( \delta x \), will choose not reject the employment offer since \( \Pi_A \) is unbounded. This implies that no arbitrageur employed by the bank will ever choose to trade more than \( \delta x/4 \) units. Hence, it is unnecessary to leave \( \overline{M} \) unbounded. Contradiction.

**Third Part:** If \( \overline{w} < \max_{x \leq \overline{M}} E[|W(X \cdot \Pi)|] \), then speculators are not screened and the bank is better off not hiring anyone. Meanwhile, if \( \overline{w} > \max_{x \leq \overline{M}} E[|W(X \cdot \Pi)|] \), the bank can hire additional arbitrators, with types just beyond \( \delta x \), at a profit without violating the speculators screening condition.\(^1\)

**Fourth Part:** The arbitrageur’s outside option at \( \Delta X_{N,t} = \overline{M}/4 \) is equal to \( 4\lambda M_L \overline{M} - 2\lambda M_L^2 \). As a result, if \( \overline{W} > 4\lambda M_L \overline{M} - 2\lambda M_L^2 \), arbitrageurs of type greater than \( \overline{M}/4 \) will choose to be employed by the bank. However, since the bank screens out speculators, these arbitrageurs could renegotiate their employment agreement with the bank after they join the firm. The bank would be willing to renegotiate because only arbitrageurs with \( \Delta X_{N,t} > \overline{M}/4 \) would attempt to do so. Understanding this, speculators would recognize that their screening condition fails if they attempt to renegotiate and would choose to accept the bank’s employment offer. In order for the contract to be renegotiation proof, the marginal arbitrageur who accepts to work for the bank must be of type \( \Delta X_{N,t} = \overline{M}/4 \) which implies that \( \overline{W} = 4\lambda M_L \overline{M} - 2\lambda M_L^2 \). Q.E.D.

**Remark on Proposition 15:** Not all arbitrageurs of type above \( M_L/4 \) and below \( \overline{M}/4 \) will choose to work for the bank.

**Proof of Proposition 16:** If \( \delta_M^* > 4\overline{M} \), explicit contracting alone does not lead to more effective arbitrage activity than the purely relational environment. Furthermore, since the bank keeps her profits from explicit contracting when she deviates on implicit promises, her IC constraint becomes:

\[
\Pi^*_B,t + V^* \geq \Pi^*_B,t + V^*_E \tag{B.9}
\]

where \( V^*_E > 0 \) is the bank’s discounted future profits from explicit contracting with arbitrageurs. In this case, the bank’s \( X_{B,t} \) is strictly higher than before and breakdown occurs earlier for a given \( V^* \). This implies that \( V^* \) is lower than in the relational environment and the limits to arbitrage problem is worsened. Q.E.D.

---

\(^1\) If the cumulative distribution function of \( F \) were not strictly monotonic, we wouldn’t be able to rule out the possibility that \( \overline{w} > \max_{x \leq \overline{M}} E[|W(X \cdot \Pi)|] \).
Bibliography


[27] Castro Rui, Clementi, Gian Luca and Glenn MacDonnalnd, 2006, ”Legal Institutions, Sectoral Heterogeneity, and Economic Development”


156
[61] Gourio, Francois, “Firms Heterogeneous Sensitivities to the Business Cycle, and the Cross-Section of Expected Returns”, working paper


[89] Lewellen, Jonathan; Stefan Nagel, and Jay Shanken, 2006, ”A Skeptical Appraisal of Asset-Pricing Tests” NBER working paper

[90] Li, Qing; Maria Vassalou, and Yuhang Xing, 2006, “Investment Growth Rates and the Cross-Section of Equity Returns,” The Journal of Business, volume 79, pages 16371665


158


[98] Malloy, Christopher; Moskowitz, Tobias J. and Vissing-Jorgensen, Annette “Long-Run Stockholder Consumption Risk and Asset Returns”, working paper


[105] Panageas, Stavros and Jianfeng Yu, 2006, ”Technological Growth, Asset Pricing, and Consumption Risk”, working paper


[119] Santos, Tano, and Pietro Veronesi, 2006b, “Habit Formation, the Cross Section of Stock Returns and the Cash Flow Risk Puzzle”, working paper


