Essays on Asset Pricing and Empirical Estimation

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Pooya Nazeran, B.S.

Graduate Program in Economics

The Ohio State University

2011

Dissertation Committee:

Pok-sang Lam, Advisor
Paul Evans
J. Huston McCulloch
© Copyright by
Pooya Nazeran
2011
Abstract

A considerable portion of the asset pricing literature considers the demand schedule for asset prices to be perfectly elastic (flat). As argued, asset prices are determined using information about future payoff distribution, as well as the discount rate; consequently, an asset would be priced independent of its available supply. Furthermore, such a flat demand curve is considered to be a consequence of the Efficient Market Hypothesis. My dissertation evaluates and questions the factuality of these assertions. I approach this problem from both an empirical and a theoretical perspective. The general argument is that asset prices do respond to supply-shocks; and changes in aggregate demand, stemming from preference changes, new international investments, or quantitative easing by the Fed, can result in price changes. Hence, asset prices are determined by both demand and supply factors. In the first essay, Downward Sloping Asset Demand: Evidence from the Treasury Bills Market, I report on my empirical study which establishes the existence of a downward sloping demand curve (DSDC) in the T-bill market. In the second essay, Asset Pricing: Inelastic Supply, I examine the theoretical issues concerning a downward sloping demand curve. I begin by clarifying a common confusion in the literature, namely, that many asset pricing models imply a flat demand curve. I show that the prominent asset pricing models, including Capital Asset Pricing Model (CAPM), Arbitrage Pricing Theory (APT) and Consumption Capital Asset Pricing Model (CCAPM), all have an underlying
DSDC. I further show that, while these models imply the relevance of supply, they are inconvenient as a vehicle for the estimation and analysis of the DSDC in the data. For those purposes, I develop an asset pricing framework based on the stochastic discount factor framework, specifically designed with a DSDC at its heart. I end the essay with a discussion of the frameworks implications and applications. In the third essay I develop on the Factor-Augmented Vector-Autoregression (FAVAR) literature, proposing a bias-corrected method. As implemented in the literature, the Principal Component Analysis stage of FAVAR introduces a classical-error-in-variable problem which leads to bias. I propose an instrument-based method for bias correction.
Vita

2003-2006 .......................... B.S. Computer Science, Mathematics
University of Maryland,
College Park, MD.

2006-2011 .......................... Graduate Teaching Associate,
Ohio State University,
Columbus, OH.

2007 ............................... Consultant,
The World Bank Treasury,
Washington, DC.

Fields of Study

Major Field: Economics

Specialization: Financial Economics, Applied Econometrics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Vita</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>1. Downward Sloping Asset Demand: Evidence from the Treasury Bills Market</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Inconsistency Between DSDC and EMH</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Alternative Interpretations</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Compatibility of EMH with DSDC</td>
<td>10</td>
</tr>
<tr>
<td>1.3.1 Future Cash Flow</td>
<td>11</td>
</tr>
<tr>
<td>1.3.2 Market Structure</td>
<td>11</td>
</tr>
<tr>
<td>1.3.3 Arbitrage Free Market</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Empirical Evidence</td>
<td>17</td>
</tr>
<tr>
<td>1.4.1 Supply Effect</td>
<td>17</td>
</tr>
<tr>
<td>1.4.2 Persistent Component</td>
<td>24</td>
</tr>
<tr>
<td>1.4.3 Structured Regression</td>
<td>27</td>
</tr>
<tr>
<td>1.5 More Evidence From the Treasury Notes Market</td>
<td>29</td>
</tr>
<tr>
<td>1.6 Concluding Discussion</td>
<td>31</td>
</tr>
<tr>
<td>2. Asset Pricing: Inelastic Supply</td>
<td>37</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>37</td>
</tr>
<tr>
<td>2.2 Literature Review</td>
<td>39</td>
</tr>
<tr>
<td>2.2.1 CAPM</td>
<td>39</td>
</tr>
<tr>
<td>Section</td>
<td>Subsection</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>2.2.2</td>
<td>APT</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Lucas Model</td>
</tr>
<tr>
<td>2.2.4</td>
<td>SDF</td>
</tr>
<tr>
<td>2.3</td>
<td>Proposed Model</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Model Description</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Implications</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Enhancements</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Empirics</td>
</tr>
<tr>
<td>2.4</td>
<td>Conclusion</td>
</tr>
<tr>
<td>3.</td>
<td>Bias Correction in FAVAR: An Instrumental-Variable Approach</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.2</td>
<td>Literature Review</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Horse Race</td>
</tr>
<tr>
<td>3.3</td>
<td>Using PCA Approximations in Regressions</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Non-Auto-Regressive Dependent Variable</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Auto-Regressive Dependent Variable</td>
</tr>
<tr>
<td>3.4</td>
<td>VAR</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Factor Augmented VAR</td>
</tr>
<tr>
<td>3.5</td>
<td>Summary</td>
</tr>
<tr>
<td>4.</td>
<td>Contributions and Future Work</td>
</tr>
</tbody>
</table>

Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Arbitrage Example</td>
<td>125</td>
</tr>
<tr>
<td>B.</td>
<td>Time Evolution of Supply-Event Response</td>
<td>127</td>
</tr>
</tbody>
</table>

Bibliography | 134
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Regression coefficients from regressing abnormal-yields on weekday and auction-day dummies</td>
</tr>
<tr>
<td>1.2</td>
<td>Sum of all T-bill auction volumes, broken down by day-of-the-week, and auction stage</td>
</tr>
<tr>
<td>1.3</td>
<td>Results from regressing on weekdays, announcement, auction, and maturity dummies</td>
</tr>
<tr>
<td>1.4</td>
<td>Regression coefficients using two different estimation windows to separate the transitory and persistent components of the response</td>
</tr>
<tr>
<td>1.5</td>
<td>Regression coefficients using two different estimation windows over the pre-April 1997 and the post-April 1997 subperiods</td>
</tr>
<tr>
<td>1.6</td>
<td>Results from both dummy-based and structured regressions</td>
</tr>
<tr>
<td>3.1</td>
<td>Bias in $Y_t^1$’s Regression Coefficients Generated Using Stock and Watson Method, its Instrumentized Variation, Bernanke et al. Method, and its Instrumentized Variation</td>
</tr>
<tr>
<td>3.2</td>
<td>Bias in $Y_t^2$’s Regression Coefficients Generated Using Stock and Watson Method, its Instrumentized Variation, Bernanke et al. Method, and its Instrumentized Variation</td>
</tr>
<tr>
<td>3.3</td>
<td>Bias in $Y_t^1$’s Regression Coefficients Generated Using Stock and Watson Method, Bernanke et al. Method, and the Proposed Method</td>
</tr>
<tr>
<td>3.4</td>
<td>Bias in $Y_t^2$’s Regression Coefficients Generated Using Stock and Watson Method, Bernanke et al. Method, and the Proposed Method</td>
</tr>
</tbody>
</table>
3.5 Forecasting Error of All Five Method; Given Time $T$ Information . . 88

3.6 RMS of deviation of IRF from Truth. . . . . . . . . . . . . . . . . . . . . . . . 95
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Abnormal-yield of 13 week maturity Treasury bills; Jan 1982 - Dec 2007</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>Left: Estimation window prior to an event. Right: Estimation window after an event</td>
<td>24</td>
</tr>
<tr>
<td>1.3</td>
<td>Unexpected component of auctions, in dollar, over 1379 auctions during the 26 year period</td>
<td>26</td>
</tr>
<tr>
<td>2.1</td>
<td>Perfectly Elastic Demand with Perfectly Inelastic Supply</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>The Effect of Positive Supply Shock on Asset Prices</td>
<td>45</td>
</tr>
<tr>
<td>2.3</td>
<td>Perfectly Inelastic Supply vs. Intra-market Supply</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>Ratio of Standard Deviation Statistic</td>
<td>71</td>
</tr>
<tr>
<td>3.2</td>
<td>Impulse Response Function for a Shock to $Y^1$: True IRF</td>
<td>89</td>
</tr>
<tr>
<td>3.3</td>
<td>Impulse Response Function for a Shock to $Y^1$: Stock and Watson Method</td>
<td>90</td>
</tr>
<tr>
<td>3.4</td>
<td>Impulse Response Function for a Shock to $Y^1$: Instrumentized Stock and Watson Method</td>
<td>91</td>
</tr>
<tr>
<td>3.5</td>
<td>Impulse Response Function for a Shock to $Y^1$: Bernanke et al. Method</td>
<td>92</td>
</tr>
<tr>
<td>3.6</td>
<td>Impulse Response Function for a Shock to $Y^1$: Instrumentized Bernanke et al. Method</td>
<td>93</td>
</tr>
<tr>
<td>3.7</td>
<td>Impulse Response Function for a Shock to $Y^1$: Proposed Method</td>
<td>94</td>
</tr>
</tbody>
</table>
3.8 Standard Errors For System of Three Equations . . . . . . . 100
3.9 Standard Deviation Ratio for System of Three Equations . . . 101
3.10 Standard Errors For System of Five Equations . . . . . . . 103
3.11 Standard Deviation Ratio Statistic for System of Five Equations . . 104
3.12 Comparison Between PCA Based Least Square and 2SLS, With Regards to Estimator Efficiency . . . . . . . . . . . . . . . . 107
3.13 Comparison Between IVLS and FGLS, in Three Equation System. Star:β, Dash:β, Dot-Dash:p . . . . . . . . . . . . . . . . . . . . 108
3.14 Coefficients from PCA-Based Regression. Dot:ρ, Dash-Dot: p, Dashed: β . . . . . . . . . . . . . . . . . . . . . . . . . . . . 113
3.15 Coefficients from 2SLS Regression. Dot:ρ, Dash-Dot: p, Dashed: β . . 114
3.16 Standard-Errors from PCA-Based Regression. Dot:ρ, Dash-Dot: p, Dashed: β . . . . . . . . . . . . . . . . . . . . . . . . 115
3.18 Standard-Deviation-Ratio Statistic from PCA-Based Regression. . . 117
3.19 Standard-Deviation-Ratio Statistic from 2SLS Regression . . . . . 118
B.1 Time evolution of the announcement coefficient . . . . . . . . 128
B.2 Time evolution of the auction coefficient . . . . . . . . . . . . . 129
B.3 Time evolution of the maturity coefficient . . . . . . . . . . . . . 130
B.4 Time evolution of the temporary component of auction coefficient . . 132
Chapter 1: DOWNWARD SLOPING ASSET DEMAND: EVIDENCE FROM THE TREASURY BILLS MARKET

If asset prices are demand elastic it may not be optimal for major market participants to instantaneously readjust their portfolios according to their information set, leading to market inefficiency. I present evidence for a downward sloping demand curve, yet argue it does not necessarily violate the efficient-market hypothesis. Furthermore, I argue the evidence is not in line with the information hypothesis, the imperfect-substitution hypothesis or the price-pressure hypothesis since it comes from the short-term Treasury bill market. The other contribution is presenting a method for isolating the persistent component of event responses, when event windows and estimation windows overlap.

1.1 Introduction

A downward sloping demand curve (DSDC) may affect the way major hedge-funds or investment-bankers readjust their portfolios. In presence of DSDC, the Securities and Exchange Commission (SEC) may have to regulate price-pegging and market stabilization practices by market makers; antitrust courts may have to consider the demand elasticity of asset prices when forcing a firm to reduce its holdings of another firm; bankruptcy courts may have to mind such demand elasticity when restructuring
a firm by selling its subsidiaries; and private equity firms buying out publicly traded firms, may examine the demand elasticity of their purchase. In corporate finance, firms trying to finance their activities through asset sales, contrary to Modigliani and Miller (1958), may have to deal with their asset’s demand elasticity, in cost of capital calculations. Last but not least, parties shorting option contracts may need to have an eye on the volume of option contracts issued, as long parties may need to go beyond Black-Scholes’ pricing and incorporate the supply of the contracts into pricing.$^1$

If asset prices can be affected by supply events that do not contain adverse information, the asset market has a DSDC. If evidence for a DSDC is found, its implications regarding market efficiency, as well as a theoretical framework for studying the aforementioned issues are of interest. This paper will empirically investigate the existence of a DSDC, as well as its implications regarding the Efficient Market Hypothesis (EMH) leaving the theoretical discussions for the accompanying paper.

There are several possibilities for what an asset market’s aggregate demand curve can look like. The first one is a perfectly elastic (flat) aggregate demand curve. With that, in a hypothetical where the market regulator imposes a price floor slightly above market equilibrium, no one would show interest in holding the asset and everyone becomes a seller, perhaps clamoring to take short positions. Alternatively individuals may have downward sloping demand curves, and in presence of many of them, the aggregate demand curve may appear virtually flat. Mathematically speaking, such an aggregate demand curve is downward sloping, but whether it can be approximated as flat or not depends on the scale of the quantity axis. The scale of the quantity axis in

$^1$In fact option traders do use Open Interest tables which show the volume of outstanding contracts for different expirations.
turn depends on the size of an individual investor looking at the aggregate demand curve. While some advocate of the flat demand curve hypothesis prescribe to the former view; some intend the second case where all individual investors’ portfolios are such that the aggregate demand curve *appears* flat to them. In other words, the former view presents a flat demand curve as a matter of principal, whereas the latter view presents it as a matter of practice. Advocates of the DSDC hypothesis support a third case - although for different reasons - positing that at least some individual investors do experience the downward slope. This paper takes this latter stance and for the reason that large investors do exist. In other words, although many investors may find the market’s aggregate demand curve flat, there are large enough participants who experience its slope; hence the slope may become relevant for asset pricing.

This disagreement pertaining the aggregate demand structure becomes particularly contentious as it may bring market efficiency into question. The efficiency of the asset market has been debated for years, and there seems to be evidence for and against it. Its proponents and opponents have been interpreting the empirical results in various ways. The information hypothesis, the price-pressure hypothesis and the imperfect-substitution hypothesis are some of the hypotheses spawned from the alternative interpretations. The many hypotheses came to avoid (or establish) a DSDC, which is perceived to violate market efficiency. This paper does not intend to get in the EMH debate, but rather intends to disentangle the two notions by establishing that the existence of a DSDC is not a statement about market efficiency.

I will start by arguing that a DSDC does not necessarily violate market efficiency; where an efficient market is defined to be one with zero risk-adjusted profit. Then I
will provide evidence from the Treasury bill (T-bill) market, arguing for the existence of a DSDC. The significance of this paper’s empirical observation is that it comes from a very liquid, default-free, fixed-income asset with frequent supply events. While previous results were much prone to interpretation in part because they were coming from risky assets, this paper’s finding can make a stronger case for a DSDC. Another significance is the theoretical explanation preceding the evidence, which makes the evidence potentially consistent with market efficiency, a perceived characteristic of the market for on-the-run T-bills.

The rest of this paper will be laid out as following. Section 1.2 will review the literature discussing the DSDC and its relation with the EMH. Section 1.3 will argue markets can be efficient in presence of a DSDC. Section 1.4 will provide empirical evidence for a supply effect in pricing T-bills. In this section, I will also show how to distinguish the permanent component of an event response from its transitory component when event windows and estimation windows overlap. Section 1.5 cites more evidence from the Treasury notes market. Finally Section 1.6 will contain concluding remarks.

1.2 Literature Review

1.2.1 Inconsistency Between DSDC and EMH

The EMH came to prominence following Fama’s seminal review (Fama, 1970) of the literature. Extensive empirical verifications of EMH led Jensen (1978) to proclaim

2High liquidity weakens the case for the price-pressure hypothesis, and weekly supply-events over many years, of an asset with perfectly certain payoff, weakens the case for the information hypothesis. Also, lack of idiocyncratic risk weakens the imperfect substitution hypothesis.
“I believe there is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Market Hypothesis”.

Fama (1970) does not make any reference to DSDC’s but Scholes (1972) writes “many authors in the theoretical literature in finance assume that a firm can regard the price of its shares, given its investment policies, as essentially independent of the number of shares it, or any shareholder, chooses to sell.” Briefly, it states that theoretical finance assumes a horizontal demand curve for assets, regardless of the scale of the quantity axis. Although there is no direct reference to market efficiency, this paper later becomes one of the major reference points for the DSDC-EMH debate.


Most papers state this without citation or proof. Shleifer (1986) however cites Modigliani and Miller (1958) stating “the home leverage idea behind the Modigliani-Miller theorem and simple cost of capital rules, obtain under the maintained assumption of horizontal demand curves for the firm’s equity.” It also cites APT and CAPM as examples of asset pricing under horizontal demand curve. Harris and Gurel (1986) cite a corporate finance textbook saying, “EMH predicts that security prices reflect all publicly available information ... [t]herefore, one corollary of the EMH is that ‘you can sell (or buy) large blocks of stock at close to the market price as long as you can convince other investors that you have no private information’.” The inner quote is from Principals of Corporate Finance, by Brealey, Mayers and Allen. The statement
still appeared in the 9th edition of the textbook, as one of six lessons learned from EMH.

Neither do Modigliani and Miller make a case for horizontal demand, nor does Brealey et al. establish the link between EMH and horizontal demand. As it seemed crucial to substantiate such a claim, inquiring for the “lesson” by Brealey et al., this paper failed to find a solid proof. However, looking at different arguments, they seem to fall into three categories. One set of arguments are made by linking an asset’s price to its future cash flow, another set are made by assuming a perfectly competitive market structure, and a final set are made by making the arbitrage-free market argument.

Cha and Lee (2001) state, “under market efficiency, equity prices should be equal to the present value of expected future cash flows. Hence, equity prices should be affected only by fundamentals such as expected cash flows and discount rates.” On the face of it, this statement rejects the effect of supply on price, just as its authors claim their “findings are consistent with a horizontal market demand curve for equities.”

For the second argument, Chen, Noronha, and Singal (2004) starts by emphasizing that “[t]he long-held assumption that stocks have perfect substitutes, and the perfect elasticity of demand that follows from it, is central to modern finance theory.” Presence of perfect substitutes may have made perfect elasticity of demand “central” to finance in line with microeconomics. Two textbook market structures in microeconomics are “Perfect Competition” and “Monopolistic Competition.” The former structure requires four conditions: First, many sellers and buyers none of whom are large in relation to total sales or purchases; second, homogeneous products; third, buyers and sellers who have all relevant information; and fourth, easy entry and exit.
There is little doubt that the stock market satisfies the first condition. For the second condition Scholes (1972) writes “[t]he shares a firm sells are not unique works of art”, which can be interpreted as saying asset shares are not “differentiated goods” sold by monopolistically competitive firms. Scholes continues the sentence by saying, “but [they are] abstract rights to an uncertain income stream for which close counterparts exist either directly or indirectly via combinations of assets of various kinds.” Indeed each asset has many shares, all of which are homogeneous rights to an uncertain income, and perhaps distributed between many shareholders that satisfy the first condition in a perfectly competitive market. Even if a few shareholders manage to monopolize on shares of an asset, according to Scholes there are other assets or portfolios that can replicate its payoff, so the second condition is still satisfied. Having the third and fourth conditions of perfect competition satisfied, we expect market participants to be price takers, and face a horizontal demand curve. In other words, the zero profit, efficient, perfectly competitive market structure has a flat demand curve, hence there is no DSDC.

On the arbitrage argument, Scholes (1972) writes, “[i]f any particular asset should be selling to yield a higher expected return due solely to the increase in the quantity of shares outstanding, this would indicate that investors expect to realize abnormal returns on this asset.” He then continues to say, “[b]ut investors seeing these profit opportunities would soon arbitrage them away.” In other words, he is saying if an asset has a DSDC, an increased supply creates an arbitrage opportunity, which then would not last long. So at best, the supply effect is transitory and interpretable as price-pressure.
With all that in mind, the first question is, if we do actually find price change in response to supply variations, have we necessarily found evidence against a flat demand curve? There is a literature disagreeing with that and Section 1.2.1.2.2 will review it. The second question is, if we do indeed find evidence for a DSDC, do we violate market efficiency? One short answer could be that perfect competition is not the only zero-profit market structure. As a counterexample, monopolistically competitive markets, relaxing the homogeneity assumption, lead to markets with DSDC, yet are zero-profit. In the asset market context, allowing for non-diversifiable idiosyncratic risk gives rise to nonhomogeneity.

What about a fixed-income asset with virtually no default-risk, hence no idiosyncratic risk? In absence of idiosyncratic risk, one can achieve DSDC by relaxing the first assumption of perfectly competitive markets. This in turn opens room for market-power. Then a forth question arises: Does such a DSDC create arbitrage opportunities? Section 1.3 will discuss these questions in great detail and Section 1.4 will present evidence for a DSDC that may be interpretable in this context.

1.2.2 Alternative Interpretations

As empirical evidence for a supply effect emerged, alternative hypotheses emerged to explain the empirical observations. One way to interpret a supply effect is the existence of a DSDC. On the other hand, if the supply event contains adverse information, a price change may simply be the result of the perfectly elastic demand curve shifting down.

A number of papers, like Shleifer (1986), Harris and Gurel (1986), and Lynch and Mendenhall (1997) examine additions to, and deletions from, market indices
Is an index inclusion or exclusion truly uninformative? The idea of information bearing supply events gave rise to the information hypothesis, pioneered by Scholes (1972). Jain (1987) as well as Dhillon and Johnson (1991) question the uninformative-ness and provide evidence against it for the S&P-500 inclusion/exclusion events. Particularly, the latter paper rejects the DSDC hypothesis in favor of the information hypothesis, claiming to uphold the EMH. On the other hand, Greenwood (2005) and Kaul, Mehrotra, and Morck (2000) argue that their respective indices of study, notably Nikkei 225 and TSE-300, indeed have had uninformative inclusion/exclusion events rejecting the information hypothesis.

Alternatively Harris and Gurel (1986) propose the price-pressure hypothesis which conjectures that at the event of a supply, some market participants must provide liquidity for the transaction, and seek rent for this service. This idea hypothesizes a short-run DSDC and a long-run perfectly elastic demand curve. In other words, it predicts that soon after a supply event there will be a price reversal in which prices revert back to their efficient full-information level. When Harris and Gurel (1986)
study the S&P-500, they find evidence for a full reversal, whereas Shleifer (1986) does not find significant evidence for a full-reversal, hence concludes a long-run DSDC does exist. Dhillon and Johnson (1991) and Jain (1987) do not find evidence for a price reversal either, while Lynch and Mendenhall (1997) and Beneish and Whaley (1996) find evidence for a partial price reversal.

Lastly, there is an imperfect-substitution hypothesis, in support of a DSDC. Wurgler and Zhuravskaya (2002) look at the inclusion/exclusion events in the S&P-500 index, but condition the effect on the existence of a perfect substitute for the asset. They observe that assets which have imperfect substitutes respond stronger to the supply event, in comparison to the assets which have a perfect or nearly perfect substitute. They hypothesize that if an asset does not have a perfect substitute, when it faces a supply shock arbitrageurs face more risk in profiting from the price differential. Consequently, they demand higher returns, which leads to a price drop. This hypothesis predicts that the greater the idiosyncratic risk of an asset is, the steeper the slope of its demand curve would be. Alternatively, assets with perfect substitutes face a flat demand curve.

In Section 1.4 I will provide evidence for the first part of the twofold argument, but before that, in Section 1.3, I argue against the second part.

1.3 Compatibility of EMH with DSDC

As mentioned in Section 1.2.1.2.1, there seem to be three general arguments, calling a DSDC in violation of EMH. First argument is that price of an asset is the present value of future cash flow; second, that asset markets are perfectly competitive; and third, that DSDC’s can create arbitrage opportunities.
1.3.1 Future Cash Flow

If an asset’s price is simply the present-value of its future cash flow, then the price must only depend on the future cash distribution, and the discount rate. Although this is typically interpreted to mean an asset’s supply has no effect on its price; within the SDF framework, the Lucas model tells us price is affected by supply through the stochastic discount factor, by which it affects the discount rate. So long as the supply of all assets is a determining factor of the discount rate, it must be in the set of relevant information for asset pricing. Therefore, for prices to reflect supply of assets is consistent with EMH. We will see in Section 1.4 how new supplies of T-bills affect the 13 week interest rate which is the inverse of the expected value of the SDF.

1.3.2 Market Structure

The second argument is pertaining to the market structure. The argument has the implicit premise that perfectly competitive markets are the only zero-profit markets. However, given easy entry and exit, and the many sellers and buyers, even without a flat demand curve, the asset market can be zero-profit. Relaxing the homogeneity assumptions, we can allow for assets to have a certain uniqueness to them. In other words, even though they are not “works of art”, if they do have non-diversifiable idiosyncratic risk, they can face a DSDC without leaving excess profit in the market. The imperfect-substitution hypothesis, as laid out towards the end of Section 1.2.1.2.2, explains it as a case of increasing the risk of arbitrageurs.

Alternatively one can relax the first condition of competitive market structure allowing for large investors. If the notion of a flat demand curve - in a perfectly

3Newly grown trees on the island, as metaphors for increased supply of assets, increase consumption and hence affect ratio of marginal utility.
competitive market - is a theoretical abstraction, it does not mean one can purchase an infinite amount and be a price-taker. Rather it means the price remains *virtually* unaffected by variations in supply, when the variations are caused by *small* market participants.

On the other hand, when price of assets are regarded “independent”\(^4\) of number of shares, or if “large blocks”\(^5\) can be sold close to market price, it means we are not talking about small participants anymore, and regardless of the scale on the quantity axis, in principal, the demand curve is flat. To understand a world where aggregate demand curve is flat, take a hypothetical market with homogeneous information, if a regulator imposes a price floor, just above market equilibrium, absolutely no one will show interest in holding the asset and everyone becomes a seller. Alternatively, allowing for heterogeneous conditioning information, even a free market fails to clear under such structure.

If the aggregate demand curve is only *virtually* flat, then it is flat *at the margin*. Indeed Modigliani and Miller may have assumed a flat demand curve only at the margin and in the interest of parsimony, or lack of sophisticated asset pricing models. A *non-marginal* market participant however, would then have to deal with a DSDC, and the empirical evidence in this paper is one of such cases. If non-marginal investors do not exist then, in practice, the demand curve is flat.

Non-marginal traders with a DSDC have the *necessary* condition for market-power, but whether they can make profit or not depends on the particular market structure. The stock market’s trading platform is price discriminating, implying

\(^4\)Refer to the quote from Scholes at beginning of Section 1.2.1.2.1
\(^5\)Refer to the quote from Brealy et al. in Section 1.2.1.2.1
allocative efficiency which induces investors to ignore demand elasticity of asset prices, hence act competitively. Since asset production is exogenous and it is then sold, purchased, or held without depreciation, the upward sloping cost schedule will be the reflection of the downward sloping revenue structure. In presence of discriminating transactions, this leads to zero profit.

Nonetheless, the mere existence of alternative zero profit market structures is sufficient for this subsection’s proposition: To present evidence for a DSDC is not sufficient for questioning market efficiency, rather one must also prove the DSDC present profitable opportunities.

### 1.3.3 Arbitrage Free Market

The arbitrage argument claims a DSDC creates profitable price disparities. I will show it does not whether perfect substitutes exist or not.

Before making my case, I will give a summary of the arbitrage based arguments. The typical argument goes as following:

- Without loss of generality, assume the existence of two perfectly substitutable assets one of which has a DSDC. If the one with DSDC, or both, experience a supply event, price of one drops, making it cheaper than its perfect substitute. Hence an arbitrageur can take a short position on the substitute, and a long on the cheaper asset, until price equality has been achieved.\(^7\)

\(^6\)This statement is assuming lack of strategic shirking in the initiation of new ventures. Issuance of bonds, options, or swaps however, is an endogenous decision. In Section 1.4 I will investigate how issuance of bills by the Treasury affects the T-bill’s price. Whether the Treasury is strategizing or not is not of interest to this paper, rather its the question of whether there is a DSDC or not, regardless of its implications to market efficiency, as the possibility exists either way.

\(^7\)The idea behind this argument is the statement by Scholes (1972) saying, “[i]f any particular asset should be selling to yield a higher expected return due solely to the increase in the quantity of shares outstanding, this would indicate that investors expect to realize abnormal returns on
• In case both previous assets face a DSDC, and are priced equally, a supply shock to one asset must reduce its price, while its perfect substitute remains at the higher price. Consequently there would be an arbitrage opportunity.

• Furthermore, an asset with a DSDC must face lower price at a supply event, even when the event is expected, which becomes all the more reason for inefficiency (profit opportunity).

First of all, one must note that arbitrage would not have worked in a world of flat demand curves, as it corrects prices by sheer creation of long and short positions. Moreover, since arbitrage trades involve shorting one asset and longing the other until the price of former is decreased and price of latter is increased to the point of equality, the supply shock to the latter has eventually led to a lower price for both assets.

A more formal statement of my rebuttal is based on two implications of arbitrage-free pricing in a linear asset pricing environment:

1- If an asset faces a price change, all portfolios that contain that asset must face a price change.

2- If the price of a portfolio changes (where portfolios are tradable as an ETN\textsuperscript{8} for example), inevitably the price of underlying assets must change, which by 1, affects the price of other portfolios that contain any of those assets.

To capture an asset's stochasticity in payoff, as well as its correlation with other assets, I will model assets as a linear combination of mean-zero independent stochastic processes, call them factors, and a constant. In this context, factors are thought of as fundamental sources of stochasticity in the asset market, and may be priced this asset.” He then continues to say, “[b]ut investors seeing these profit opportunities would soon arbitrage them away.” This statement was quoted in Section 1.2.1.2.1.

\textsuperscript{8}Exchange Traded Notes.
differently. In particular the Lucas model positively prices the factor which represents the ratio of marginal utility of consumption, and prices all the other factors zero.

The underlying factors can be thought of as virtual assets, where the actual assets become a portfolio of these virtual assets. In a linear asset pricing model, underlying factors must be priced consistent with actual assets; such that in line with the implications above, if a factor faces a price change, all portfolios containing that factor face a price change accordingly.

Assume an asset with DSDC experiences a supply shock. The shock gets transmitted to the asset’s underlying factors, changing their overall supply in the market, and affecting their price by implication 2. Then by implication 1, the price change is transmitted to other assets with loadings on those factors. The shock ripples throughout the market affecting the price of many assets, until a new equilibrium is reached.\(^9\)

Fundamentally the argument hinges on thinking of assets as portfolios of factors and thinking of a supply shock to an asset, as a supply shock to the underlying factors. A clarifying example has been appended to this paper.

At the supply event, an asset faces a price change, but until the substitution effect is fully worked out, the market is in a temporary disequilibrium. So the price change will be followed by a partial price reversal, as the substitution effect kicks in through the shared factors. The magnitude of price reversal has to do with how many substitutes exist, how perfect of a substitute they are, and how liquid they

\(^9\)Alternatively, take an informative news shock since it is more familiar. When a news shock affects an asset’s price, it could be through a readjustment of the factor loadings (portfolio weights), or the factor prices. In the latter case, as the price of the affected factors change, so does the price of the asset. But consequently the price of other assets depending on the affected factors must change as well. If the asset pricing function consistently prices assets and their portfolios, it will consistently price factors and their portfolios (actual assets) such that it will leave no profitable price disparities in the market.
are, consistent with the findings of Wurgler and Zhuravskaya (2002). Hess and Frost (1982) fail to reject a horizontal demand curve in part because they look at supply of assets not supply of factors, hence fail to account for substitution effects.

Observationally, this substitution effect will be indistinguishable from a price-pressure effect, and becomes a matter of interpretation. Given the coherence of the DSDC story, the evidence for it, and its potential consistency with market efficiency, this paper adopts the DSDC interpretation for the empirical observation in the following section.

Lastly, the above argument holds whether perfect substitutes exist or not; but it also holds whether the supply event is expected or not. The news of a supply event in the future is news about a new market equilibrium. The market will then have to move on an arbitrage-free trajectory, from its current state to the future equilibrium, such that it would reach the new equilibrium in time. It is noteworthy that if the news of a supply shock comes before the event, a price change should only be observable at the announcement time and not at the actual supply time. Along the same line, Hess and Frost (1982) states that when “new issues are announced well before the issue day, in an efficient market there will be no price effects of new issues around the issue day”. Moreover, if the announcement comes early enough, a price jump may not even occur at announcement time, but rather a smooth transition from current equilibrium to future equilibrium will happen. If there is uncertainty regarding the supply volume however, such that the risk-neutral expectation of new supply is different from the actual supply amount, a price change at supply time may still occur.
1.4 Empirical Evidence

Section 1.4.1.4.1 will provide evidence for a supply effect which at least testifies to a transitory effect. Section 1.4.1.4.2 will then propose and execute a method for validating the existence of a persistent component in the event response. Section 1.4.1.4.3 imposes a structure on the econometric analysis. Lastly, a study of the time evolution of the event responses has been appended to this paper.

1.4.1 Supply Effect

This section will conduct an event study on T-bills with 13 week maturity (13W), chosen in part due to their liquid market.\textsuperscript{10} The data is daily and obtained from Federal Reserve’s constant maturity\textsuperscript{11} TCMNOM series, starting with the first business day of 1982, through the last day of 2007.\textsuperscript{12} The event window is only the day of the event. Since we are looking at bills, instead of normal-return and abnormal-return - in the terminology of MacKinlay (1997) - we would be looking at normal-yield and abnormal-yield. Given the object of study is zero-premium, a constant mean return model is suitable for estimation of the normal-yield. The estimation window will be

\textsuperscript{10}Reopens with maturities between 8 to 19 weeks, being close substitutes, were also included.

\textsuperscript{11}The Fed publishes constant maturity yields by interpolating most recently auctioned bills into a yield-curve. The advantage of using these interpolated yields, rather than actually traded bills is that they more accurately reflect the interest rate. Off-the-run bills are less favorable in repo markets, hence they are less liquid and so have higher yields.

\textsuperscript{12}From Jan 4th 1982 through Dec 31st 2007. 2008 and 2009 data were left out as the temporary disruption in credit markets could have biased the results.
Figure 1.1: Abnormal-yield of 13 week maturity Treasury bills; Jan 1982 - Dec 2007

the 4 days prior to a day of interest. In mathematical terms:

\[ NY_t = \frac{Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4}}{4} \]  

(1.1)

\[ ABY_t = Y_t - NY_t \]  

(1.2)

where \( Y_t \) is the yield at time \( t \), \( NY_t \) is the normal-yield at time \( t \), and \( ABY_t \) is the

abnormal yield at time \( t \). Figure 1.1 helps get a feeling for the abnormal-yields.

Once the abnormal-yields are estimated, a GARCH(3,1) is performed with Generalized Error (GED) residuals over twenty-one autoregressive orders, to account for the residuals’ heteroskedasticity, leptokurticity as well as autocorrelation, respectively.

\(^{13}\)The estimation window may not be the four observations preceding the date in the time-series, rather it is defined over business days. Yet, if one of those business days happens to be a holiday, or for another reason data is not available, its yield is estimated to be the average yield of the rest of the estimation window. This is important to make the day-of-the-week estimations meaningful.

\(^{14}\)The abnormal-yield has a mean of -0.00337, a standard-deviation of 0.1169, and after normalization, a skewness of -1.886, and kurtosis of 38.48.
Our first regression simply studies the day-of-the-week seasonality effect in the following setup:

\[ ABY_t = \beta^M D_{t}^{\text{Mon}} + \beta^T D_{t}^{\text{Tue}} + \beta^W D_{t}^{\text{Wed}} + \beta^R D_{t}^{\text{Thu}} + \beta^F D_{t}^{\text{Fri}} + \epsilon_t \]  

(1.3)

where \( D_{t}^{\text{Monday}} \) is a dummy variable for day \( t \) being a Monday. The coefficients with their t-values are exhibited in the first line of Table 1.1. One would expect the coefficients to be statistically insignificant, but to the contrary Monday and Tuesday dummies turn out to have significantly positive coefficients, while Thursday and Friday dummies have significantly negative coefficients. In an attempt to explain some of the day-of-the-week effect, the second line of Table 1.1 introduces an auction-day dummy for 13W bills.\(^{15}\)

\[ ABY_t = \beta^M D_{t}^{\text{Mon}} + \beta^T D_{t}^{\text{Tue}} + \beta^W D_{t}^{\text{Wed}} + \beta^R D_{t}^{\text{Thu}} + \beta^F D_{t}^{\text{Fri}} + \beta^{\text{Auc}} D_{t}^{\text{Auc}} + \epsilon_t \]  

(1.4)

By adding a dummy for auction days we verify a yield increase upon an auction.\(^{16}\) Moreover, upon the introduction of the auction-dummy, the Monday effect shrinks significantly and the Tuesday effect weakens while the rest of the day-of-the-week effects remain statistically intact. To understand why, one needs to understand the auctioning mechanism for T-bills. The Treasury first makes an announcement on “announcement date”, indicating an “auction date”, the amount of bills being “offered”, and the “maturity date”. The offering amount is dollar denominated, and called the “offer”. The announcement also indicates the “issuing date”, which is when the bills

\(^{15}\)Table data are all Garch(3,1) on GED errors. GLS regressions with time-varying variance estimated over a 60 day windows produces similar results. Coefficients and standard-errors were slightly different but statistical significance was the same.

\(^{16}\)This paper hypothesizes that increased supply corresponds to lower price. In case of bonds, price moves in the opposite direction of yield.
Table 1.1: Regression coefficients from regressing abnormal-yields on weekday and auction-day dummies

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Auction Day</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0164</td>
<td>0.0096</td>
<td>-0.0013</td>
<td>-0.0063</td>
<td>-0.0069</td>
<td></td>
<td>0.528</td>
</tr>
<tr>
<td>(13.7557)</td>
<td>(8.1731)</td>
<td>(-1.0850)</td>
<td>(-5.3160)</td>
<td>(-5.8812)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0053</td>
<td>0.0079</td>
<td>-0.0015</td>
<td>-0.0060</td>
<td>-0.0064</td>
<td>0.0107</td>
<td>0.530</td>
</tr>
<tr>
<td>(2.6200)</td>
<td>(6.6499)</td>
<td>(-1.3180)</td>
<td>(-5.1521)</td>
<td>(-5.5643)</td>
<td>(6.7649)</td>
<td></td>
</tr>
</tbody>
</table>

Garch(3,1) on Generalized Errors (GED) with 21 autoregressive order. T-values inside parenthesis. Daily data from Jan 1982 through Dec 2007. The top regression’s residuals’ excess kurtosis is 6.52, while that of the lower one is 6.63. The estimated GED parameter is 1.045 for the top regression and 1.052 for the bottom regression, both statistically significantly non-normal. GED parameter of 1.05 yields excess-kurtosis of around 2.5.

will be “issued” to the winners and is typically two to three days after the auction. The auction is held on the auction day, which is usually three to four days after the announcement. The auction may sell more or less than the “offered” amount. The amount actually sold is called the “issued” amount, and the difference between offered and issued will be called the “unexpected” amount in this paper. The auction result is announced shortly after the auction, on the same day.

Table 1.2 helps with understanding the volume of activity on any day of the week. Starting with the first column from left, the first row discloses the total offered amount of all 13W auctions which were announced on a Monday to be $371B. The second line tells us the offered amount of all auctions announced on a Tuesday, from Jan 1982 until Dec 2007, add up to $7.9T. Overall, the first column states that, by offered amount, most auctions were announced on Tuesdays and Thursdays. The second and third columns state whether we look at the issued amount or unexpected amount, most auctions were held on Mondays, and after that, on Tuesdays. The fourth and fifth columns show most bills were issued and matured on Thursdays.
Table 1.2: Sum of all T-bill auction volumes, broken down by day-of-the-week, and auction stage

<table>
<thead>
<tr>
<th>Weekday</th>
<th>Announcement by Offer</th>
<th>Auction date by Issue</th>
<th>Auction date by Unexpected</th>
<th>Issuance date by Issue</th>
<th>Maturity date by Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>371.25</td>
<td>16712.88</td>
<td>2754.63</td>
<td>256.90</td>
<td>0</td>
</tr>
<tr>
<td>Tuesday</td>
<td>7946.35</td>
<td>2336.41</td>
<td>381.01</td>
<td>101.23</td>
<td>30.00</td>
</tr>
<tr>
<td>Wednesday</td>
<td>424.75</td>
<td>337.05</td>
<td>10.85</td>
<td>62.21</td>
<td>0</td>
</tr>
<tr>
<td>Thursday</td>
<td>7930.00</td>
<td>302.63</td>
<td>1.93</td>
<td>18714.68</td>
<td>19128.82</td>
</tr>
<tr>
<td>Friday</td>
<td>91.30</td>
<td>223.24</td>
<td>0.14</td>
<td>777.20</td>
<td>753.40</td>
</tr>
</tbody>
</table>

In billions of dollars. Aggregate dollar volume from Jan 1982 through Dec 2007. T-bill auctions were generally announced on Tuesdays and Thursdays, and held on Mondays. T-bills were generally issued on Thursdays, and matured on Thursdays as well.
The second and third columns of Table 1.2 help with understanding the regression results from Table 1.1. Since most auctions are held on Mondays and after that on Tuesdays, an auction-day dummy accounts for most of the Monday effect and some of the Tuesday effect.

With the insight given by Table 1.2, the next regression to try should include an announcement-day dummy or a maturity-day dummy. Table 1.3 shows regression results. Note that at announcement-day we are observing the yield of less-than-a-week old 13W bills respond to learning about their new siblings being born in a few days.

To simplify comparison, the first line repeats results from Table 1.1. The second line says an announcement effect is borderline significant and quite weak, at one fifth of a basis-point. The announcement dummy also leaves a small dent in the Tuesday and Thursday effects. A maturity day dummy however, in the third line, virtually eliminates the Thursday effect. At maturity the yield of less-than-a-week old 13W bills responds to the maturity of their 13W old siblings. A maturity event is a negative supply shock; and consistently, the regression results show that yields tend to drop on maturity days. The forth line summarizes this section, by including an announcement, auction and maturity dummy together. Knowing there is no Thursday effect, once maturities are accounted for, the elimination of the Thursday dummy does not cause an endogenity problem, but actually fixes a near-multi-colinearity problem. Now it is evident that the maturity event has a significant effect and inflicts a three-quarter of a basis point drop to the yield.\footnote{An abnormal-yield of 1.09bps per auction is \emph{sine} the day-of-the-week, meaning in reality, if the auction is held on Monday, a Monday abnormal-yield must also be added. Likewise for auctions which are held on other days. In this regard, announcements, frequently occurring on Tuesdays, are going to have a larger observable abnormal-yield, meaning an announcement abnormal-yield \emph{cum} sine day-of-the-week.}
Table 1.3: Results from regressing on weekdays, announcement, auction, and maturity dummies

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Announcement</th>
<th>Auction Day</th>
<th>Maturity Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0164</td>
<td>0.0096</td>
<td>-0.0013</td>
<td>-0.0063</td>
<td>-0.0069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13.7557)</td>
<td>(8.1731)</td>
<td>(-1.0850)</td>
<td>(-5.3160)</td>
<td>(-5.8812)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0164</td>
<td>0.0085</td>
<td>-0.0015</td>
<td>-0.0076</td>
<td>-0.0067</td>
<td>0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14.0274)</td>
<td>(7.0435)</td>
<td>(-1.2734)</td>
<td>(-5.9098)</td>
<td>(-5.8640)</td>
<td>(2.3849)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0164</td>
<td>0.0094</td>
<td>-0.0015</td>
<td>-0.0014</td>
<td>-0.0064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13.8906)</td>
<td>(8.0840)</td>
<td>(-1.2884)</td>
<td>(-0.4598)</td>
<td>(-5.4561)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0053</td>
<td>0.0065</td>
<td>-0.0018</td>
<td>-0.0057</td>
<td>0.0024</td>
<td>0.0109</td>
<td>-0.0073</td>
<td></td>
</tr>
<tr>
<td>(2.6237)</td>
<td>(5.3568)</td>
<td>(-1.6007)</td>
<td>(-5.3244)</td>
<td>(2.7682)</td>
<td>(6.9202)</td>
<td>(-6.4280)</td>
<td></td>
</tr>
</tbody>
</table>

Garch(3,1) on Generalized Errors (GED) with 21 autoregressive order. T-values inside parenthesis. Daily data from Jan 1982 through Dec 2007. The last regression’s residuals’ excess kurtosis is 6.67, and the estimated GED parameter is 1.048, yielding about 2.5 excess kurtosis.
1.4.2 Persistent Component

A common method used in checking for persistence is the cumulative-abnormal-response (CAR). This method however, does not work when event windows and estimation windows overlap. Since T-bill related events are very frequent, CAR cannot be applied. The day after a 13W auction, there usually is an auction for a close substitute, namely the 4W bill, and two days later, old-old off-the-run 13W bills mature. To identify the persistent component of an event response, the event response for all events must be simultaneously accounted for.

This will be achieved by shifting the estimation window. Figure 1.2 visualizes the approach. The schematic on the left shows that when the estimation window is prior to the event window, the abnormal-yield is composed of both the transitory and persistent components of the response. The schematic on the right however, shows when the estimation window follows the event window the abnormal-yield only embodies the transitory component.

its day-of-the-week effect can be stronger. The observable abnormal-yield is important for studying the event’s effect on trading strategies.

Figure 1.2: Left: Estimation window prior to an event. Right: Estimation window after an event
Table 1.4 repeats the regression in the last line of Table 1.3 but with the two different estimation windows. The columns of interest are the three on the right. Upon an announcement there is a quarter of a basis point yield change. The transitory component is half, at 0.0013, but statistically insignificant. On an auction day, the yield increases by about 1.09 bps but half of that is transitory, while the other half carries on. For maturity however, the whole abnormal-yield is transitory.

Given the discussion at the end of Section 1.3, a small abnormal-yield at announcement time is not surprising, but the observations regarding the maturity and auction events may be at odds with market efficiency. On the surface of it, the supply effect at maturity is just like that of an auction but negative. Consequently, one would expect the transitory and persistent components of maturity to be negative of that of an auction; however, since the event of maturity is announced over 13 weeks in advance, the market has reached the new equilibrium and the lack of a persistent price change is expected. The transitory overshoot though, can be interpreted as temporary disequilibrium, both in case of maturity and auction events. The unexpected observation here, is the existence of a permanent component at the pre-announced auction event. An explanation can be uncertainty towards the supply size, due to the difference between the offered amount and issued amount.

Figure 1.3 shows the unexpected component of each auction’s supply, notably the difference between the dollar amount actually issued, and the dollar amount announced. There have been 1379 auctions over the 26 year period, and evidently, beginning April 1997 there has been a regime change. Indeed the standard deviation of the unexpected component, prior to April 1997 is $2.61e + 8$ and if we exclude the one anomalously large observation before April 1997, the standard deviations becomes
Figure 1.3: Unexpected component of auctions, in dollar, over 1379 auctions during the 26 year period.

1.65e + 8. In the post-April 1997 period, standard deviation is 1.59e + 9, which is about 10 times larger.

Top two lines of Table 1.5 show pre-April 1997 data and bottom two lines show post-April 1997 data and clearly during this post-April 1997 period, the response to an announcement event has become insignificant. Moreover the permanent component of an auction has shrunk during recent and more uncertain period. One explanation can be that the risk-neutral expectation of new supply has been closer to the actual amount. However, given the volatility, this is not a compelling explanation.

There may be another event around mid-1997 biasing our study. Nonetheless, we are left with a puzzling situation in which there is a shift away from announcement event response and towards stronger yet more transitory auction event responses in more recent periods. Using a time evolution study Appendix.B sheds some light on this issue.
To summarize, this section shows that these events have both a permanent and a transitory effect. Whether these effects point to inefficiency or not is a non-trivial question as the latter discussion in this subsection tries to show. Nevertheless, what we can take away is the existence of statistically meaningful and permanent abnormal-yields, regardless of their implication to market efficiency.

1.4.3 Structured Regression

Up to this point, the empirical work is pretty conventional. A standard event-study is executed on a number of events, represented by dummy regressors. One would however, like to go beyond a dummy based regression and impose structure to study more specifically how the volume of an auction affects the yield. One can arbitrarily throw in auction volume as a regressor, but given its non-stationarity such naive attempt will fail. One may choose to detrend auction volume, apply the HP filter, or use auction-volume to outstanding-bills ratio, but these all make ad-hoc regression equations.

I will use Nazeran (2010a) as the theoretical guide for the empirical work in this section, in particular the following equation from the implications sections:

$$R_{f,t} = \frac{\sum_{a=1}^{A} Q_{a,t} \cdot (E[P_{a,t+1}] - i_{a,t}^n(\cdot))}{\sum_{n=1}^{N} W_t^n}$$

(1.5)

The equation has a very simple derivation: Let $P_{a,t}$ be the price of asset $a$ at time $t$. Let $Q_{a,t}$ be the number of outstanding shares of asset $a$, at time $t$. Let $W_t^n$ be the wealth of investor $n$ at time $t$. Then we have an identity equating the total wealth in the market with the value of all assets: $\sum_a Q_{a,t} P_{a,t} = \sum_n W_t^n$. The cited paper develops a framework which results in the following equation for asset prices: $P_{a,t} R_{f,t} = E[P_{a,t+1}] - i_{a,t}^n(\cdot)$. The right hand side is the expected future payoff minus marginal risk-premium. The structure is quite similar to mainstream asset pricing models. It is the specification of $i(\cdot)$ which makes this asset pricing framework special, but for the purpose of T-bills, $i(\cdot)$ is zero. That is because $i(\cdot)$ is the premium incurred due to uncertainties in the future payoff, $P_{a,t+1}$, and a function of its centered-moments. For default-free bonds however, this value is zero. Plugging the second equation into the first gives us equation 5.
The left-hand-side is the risk-free rate of return at time $t$. The numerator on the right hand side is a value-weighted summation, over all assets, of $E[P_{a,t+1}] - i^n_a(.)$ which is expected future payoff minus marginal risk premium. The weight $Q_a$ is the quantity of outstanding shares of asset $a$. In the denominator sits the total wealth in the asset market. Knowing $i^n_a(.) = 0$ for fixed-income assets, let $NR_{f,t}$ be the normal risk-free rate, as defined earlier in this paper, then the abnormal risk-free rate is:

$$ R_{f,t} - NR_{f,t} = \frac{\Delta Q_{tbill,t} \cdot P_{tbill,t+1} + \epsilon_t}{\sum_{n=1}^{N} W^n_t} $$

(1.6)

The numerator has two terms, the first one due to a supply-change in T-bills and the second one due to all other changes in the market stemming from change in expected payoffs or risk of other assets. $\Delta Q_{tbill,t} \cdot P_{tbill,t+1}$ is simply the face-value (FV) of auctioned bills. Assuming $\epsilon_t$ is uncorrelated with T-bill supply changes, we need a proxy for $\sum_{n=1}^{N} W^n_t$ to run a regression. We will assume $\sum_{n=1}^{N} W^n_t$ is proportional to CRSP’s value-weighted index including dividend (vwindd), naming the constant of proportionality $\alpha$.

$$(R_{f,t} - NR_{f,t}) \cdot vwindd = \frac{FV_{Announced}}{\alpha} + \frac{FV_{Auctioned}}{\alpha} + \frac{FV_{Matured}}{\alpha} + \frac{\epsilon_t}{\alpha}$$

(1.7)

The regression results appear in Table 1.6. The upper-half is repeating results from Table 1.4, and the bottom-half is reporting results from the regression equation above. Standard-errors are also added inside square-brackets. Day-of-the-week dummies are still included to partially account for what appears as $\epsilon_t$ in the equation above.

The structure considerably improves announcement and auction t-values. Using volume data also seems to explain away the Tuesday effect. The dummy regression tells us an auction increases the yield by 1.09 bps, but the structured regression is more difficult to interpret. There we learn for every $1B of 13W T-bills auctioned,
the abnormal-yield $vwindd$ product increases by 1.4429. If we could assume an auction’s effect on $vwindd$ is negligible,$^{19}$ the result states that for every $1B auctioned, the yield changes by $\frac{1.4429}{vwindd}$. Further, we learn that an announcement’s effect on abnormal-yield is one-forth that of an auction, and a maturity’s effect on abnormal-yield is one-third that of an auction.

Note the equation above is implying that all event regressors will have the same coefficient, notably $\frac{1}{\alpha}$, but this does not materialize because the nature of the events are different. The announcement event is news of a future equilibrium, the auction event brings about the new equilibrium and the maturity event is bringing about a long-anticipated new equilibrium so each has different levels of effect; but the imposed structure is not sophisticated enough to account for those difference. In fairness to the structure however, such a small standard-error for the auction coefficient is testimony to equation 5’s ability to explain risk-free rate of return.$^{21}$

Table 1.6 gives us further confidence that announcements and auctions increase the yield, and maturities reduce them. It also tells us the effect on yield is increasing in volume.

1.5 More Evidence From the Treasury Notes Market

In its March 18th FOMC press release (FOMC, 2009), the Fed announced its intention “... to purchase up to $300 billion of long-term Treasury securities ... .” This announcement was followed by a 51 bps drop in the 10-year note, a 46 bps

$^{19}$Results not reported in this paper support this assumption.

$^{20}$For example, Oct 17th 1996 when $vwindd$ was 1443, a $10B auction would have increased the yield by 1bps.

$^{21}$It is noteworthy that an equation that links risk-free rate of return to equity risk and return is an ongoing challenge in the literature, in part through the channel of the Equity-Premium Puzzle.
drop in the 5-year note, and a 23 bps drop in the 2-year note yield. This news of a 
negative supply shock could have increased the price only in presence of a forward 
looking investor with a downward sloping demand curve.\(^2\) Ten business days later, 
the 10-year bond had gained 17 bps, the 5-year note had gained 11 bps and the 2-year 
note was almost unchanged, indicating most of the shock was permanent. 

This may be a supply event pointing at a DSDC, but one may apply the informa-
tion hypothesis on the basis that such liquidity injection contains adverse information 
about future inflation. If this was the case, the yield on inflation-indexed notes should 
not have changed,\(^3\) however, we see the yield on 10-year TIPS dropped by 62 bps 
and the yield on 5-year TIPS dropped by 47 bps. Ten business days later, the 10-year 
TIPS had gained 9 bps and the 5-year TIPS was virtually unchanged. The conclusion 
is, even if the Fed’s announcement was not completely uninformative about inflation, 
at least the bulk of the yield change was not due to a change in inflation expectations, 
hence one cannot reject the DSDC hypothesis. 

Alternatively, one may claim the announcement did not contain adverse informa-
tion about future inflation, but rather it contained adverse information about the 
state of the economy. It is noteworthy that this announcement came six months after 
the Bear Stearns, AIG, Lehman fiasco, 19 days after 08Q4 GDP data was released 
and 12 days after Feb-09 employment data was released. With all that in mind, it is 
fair to say the market was already well aware of the state of the economy, and had 
priced it in. Lastly, the Dow Jones Industrial Average index closed at 7395.70 on 
March 17th, 7486.58 on Mach 18th, 7400.80 on March 19th, indicating no abnormal

\(^2\)The Fed technically has a perfectly inelastic demand curve, but it is easier to think of the event as a negative supply shock

\(^3\)A nominal bond’s yield is inflation plus real-interest-rate, \(y_t = i_t + r_t\). Information about higher inflation must increase the spread between the two rates of return.
feelings about the equity market. Likewise, the more broadly defined S&P-500 index closed at 778.12 on March 17th, 794.35 on March 18th, 784.04 on March 19th.

Although the Fed did not announce how the $300B is to be distributed between nominal and inflation-indexed bonds, a quick review of the Fed’s balance-sheets shows, as of October 2009, 98% of it had been spent on nominal notes.24

1.6 Concluding Discussion

This paper presents empirical evidence for an aggregate downward sloping demand curve (DSDC), even though the efficacy of “arbitraging” in itself is sufficient evidence for an aggregate DSDC. This paper also presents a number of arguments for why a DSDC is not sufficient to violate the Efficient Market Hypothesis (EMH). For instance, the paper shows a consistent linear asset pricing model can allow for asset prices to respond to supply, while not leaving profitable price disparities in the market. More specifically, it can price them as discounted future cash flows, while allowing for prices to respond to supply events, through the discount rate.

To observe that T-bill prices are in part determined by their supply should not be surprising. Open market operation involves affecting the interest rate by manipulating the supply of assets, usually short-term government securities. If it was not for a DSDC in the asset market, the Open Market Desk at the Federal Reserve Bank of New York would not have been able to keep Fed Funds rate close to the Target rate.

Numerous cases have been made for the existence of a DSDC. In perceived defense of EMH, alternative interpretations for the empirical observations have been made.

24On a side note, the response by the TIPS is an example of how substitutes, sharing factors with a shocked asset, face a price change. Moreover, the same-day response of TIPS’s yield points to the liquidity of the government securities market and the little room there is for rent-seeking as hypothesized by the price-pressure hypothesis.
trying to bring the existence of a DSDC into question. This paper presents evidence from the 13 week US Treasury bill, a very liquid asset, with no income uncertainty, new supplies of which are auctioned every week. These weekly supplies have been going on for almost 30 years, and given the US credit rating, one can not argue auctioning conveys unexpected adverse information. On the other hand, the transitory component of both auction and maturity events may be due to market-makers demanding excess-return for their service. This would then support the price-pressure hypothesis which states that upon supply perturbation market makers provide liquidity (or asset) and demand excess-return. This dynamic however occurs in illiquid or somewhat liquid markets, not markets like that of the T-bill. Through arguments in Section 1.3.1.3.3, this paper argues that the price change following a supply perturbation ignites a cascade of substitution effects, which once worked out, ameliorate some of the supply effect. Specifically, the evidence in this paper affects the discount rate, which clearly affects other assets’ value.

All in all, the empirical observation in this paper makes a strong case for the existence of a DSDC in asset markets. This does not necessarily imply inefficiency, as argued in the first half of the paper. But it does note concerned parties that the practically flat demand curve may become downward sloped in certain investment practices.

This paper has three major contributions. The first one is to argue the existence of an aggregate DSDC is not sufficient to reject the EMH. The second one is to show the T-bill market responds to positive and negative supply events. And lastly, to present a new method for isolating the persistent component of event responses, when event
windows and estimation windows overlap.
Table 1.4: Regression coefficients using two different estimation windows to separate the transitory and persistent components of the response

<table>
<thead>
<tr>
<th>Estimation Window</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Friday</th>
<th>Announcement</th>
<th>Auction</th>
<th>Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Event</td>
<td>0.0053</td>
<td>0.0065</td>
<td>-0.0018</td>
<td>-0.0057</td>
<td>0.0024</td>
<td>0.0109</td>
<td>-0.0073</td>
</tr>
<tr>
<td></td>
<td>(2.6237)</td>
<td>(5.3568)</td>
<td>(-1.6007)</td>
<td>(-5.3244)</td>
<td>(2.7682)</td>
<td>(6.9202)</td>
<td>(-6.4280)</td>
</tr>
<tr>
<td>After Event</td>
<td>0.0091</td>
<td>0.0056</td>
<td>-0.0041</td>
<td>-0.0065</td>
<td>0.0013</td>
<td>0.0056</td>
<td>-0.0086</td>
</tr>
<tr>
<td></td>
<td>(4.0622)</td>
<td>(4.0597)</td>
<td>(-3.1262)</td>
<td>(-5.3054)</td>
<td>(1.3178)</td>
<td>(3.2375)</td>
<td>(-6.6505)</td>
</tr>
</tbody>
</table>

Table 1.5: Regression coefficients using two different estimation windows over the pre-April 1997 and the post-April 1997 subperiods.

<table>
<thead>
<tr>
<th>Duration</th>
<th>Est. Window</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Friday</th>
<th>Announcement</th>
<th>Auction</th>
<th>Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 82</td>
<td>Before Event</td>
<td>0.0022</td>
<td>-0.0188</td>
<td>-0.0038</td>
<td>-0.0116</td>
<td>0.0259</td>
<td>0.0099</td>
<td>-0.0114</td>
</tr>
<tr>
<td></td>
<td>Perm.+Trans.</td>
<td>(0.6401)</td>
<td>(-2.5052)</td>
<td>(-1.9087)</td>
<td>(-6.1580)</td>
<td>(3.5985)</td>
<td>(3.6136)</td>
<td>(-6.1231)</td>
</tr>
<tr>
<td>Mar 97</td>
<td>After Event</td>
<td>0.0153</td>
<td>-0.0070</td>
<td>-0.0011</td>
<td>-0.0073</td>
<td>0.0195</td>
<td>0.0021</td>
<td>-0.0086</td>
</tr>
<tr>
<td></td>
<td>Transitory</td>
<td>(3.7807)</td>
<td>(-0.7085)</td>
<td>(-0.4978)</td>
<td>(-3.4023)</td>
<td>(2.0470)</td>
<td>(0.6589)</td>
<td>(-3.9241)</td>
</tr>
<tr>
<td>Apr 97</td>
<td>Before Event</td>
<td>0.0053</td>
<td>0.0068</td>
<td>-0.0011</td>
<td>-0.0042</td>
<td>0.0009</td>
<td>0.0113</td>
<td>-0.0049</td>
</tr>
<tr>
<td></td>
<td>Perm.+Trans.</td>
<td>(2.1256)</td>
<td>(4.5867)</td>
<td>(-0.7733)</td>
<td>(-3.0241)</td>
<td>(0.6178)</td>
<td>(5.7697)</td>
<td>(-2.9795)</td>
</tr>
<tr>
<td>Dec 07</td>
<td>After Event</td>
<td>0.0056</td>
<td>0.0034</td>
<td>-0.0049</td>
<td>-0.0065</td>
<td>-0.0008</td>
<td>0.0074</td>
<td>-0.0068</td>
</tr>
<tr>
<td></td>
<td>Transitory</td>
<td>(2.0745)</td>
<td>(2.0240)</td>
<td>(3.0474)</td>
<td>(-4.1071)</td>
<td>(-0.5012)</td>
<td>(3.5932)</td>
<td>(-3.8019)</td>
</tr>
</tbody>
</table>

Garch(3,1) on Generalized Errors (GED) with 21 autoregressive order. T-values inside parenthesis. “Perm.+Trans.” reminds that when the estimation window is before the event, the event-response is the summation of permanent response and transitory response; however, when estimation window is after the event, the event-response only embodies the transitory response.
Table 1.6: Results from both dummy-based and structured regressions.

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Friday</th>
<th>Announcement</th>
<th>Auction</th>
<th>Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dummy</td>
<td>0.0053</td>
<td>0.0065</td>
<td>-0.0018</td>
<td>-0.0057</td>
<td>0.0024</td>
<td>0.0109</td>
<td>-0.0073</td>
</tr>
<tr>
<td></td>
<td>[0.0020]</td>
<td>[0.0012]</td>
<td>[0.0011]</td>
<td>[0.0011]</td>
<td>[0.0009]</td>
<td>[0.0016]</td>
<td>[0.0011]</td>
</tr>
<tr>
<td></td>
<td>(2.6237)</td>
<td>(5.3568)</td>
<td>(-1.6007)</td>
<td>(-5.3244)</td>
<td>(2.7682)</td>
<td>(6.9202)</td>
<td>(-6.4280)</td>
</tr>
<tr>
<td>Structured</td>
<td>-4.9342</td>
<td>1.4748</td>
<td>-1.6631</td>
<td>-5.3619</td>
<td>0.3747</td>
<td>1.4429</td>
<td>-0.5423</td>
</tr>
<tr>
<td></td>
<td>[1.5398]</td>
<td>[1.3797]</td>
<td>[1.0650]</td>
<td>[1.0087]</td>
<td>[0.0904]</td>
<td>[0.0908]</td>
<td>[0.0885]</td>
</tr>
<tr>
<td></td>
<td>(-3.2044)</td>
<td>(1.0689)</td>
<td>(-1.5617)</td>
<td>(-5.3155)</td>
<td>(4.1455)</td>
<td>(15.8900)</td>
<td>(-6.1304)</td>
</tr>
</tbody>
</table>

Garch(3,1) on Generalized Errors (GED) with 21 autoregressive order. GED coefficient of 1.065 for the second regression. T-values inside parenthesis and standard-errors inside square brackets. Daily data from Jan 1982 through Dec 2007. The amount used for announcement is the offered amount, whereas the amount used for auction, as well as maturity, are the actual issued amounts. In the structured regression, day-of-the-week regressors are dummies, while the announcement, auction and maturity regressors are volume, in billions of dollars.
Chapter 2: ASSET PRICING: INELASTIC SUPPLY

There is empirical evidence suggesting the existence of a downward sloping demand curve in the asset market. Little work however has been done on the theory side. This paper examines the popular asset pricing models, explaining why a new framework is necessary to study downward sloping demand curves. Lastly, a new framework will be proposed.

2.1 Introduction

The existence of downward sloping demand curves (DSDC), and the resulting demand elasticity of asset prices, can affect both the pricing and the trading of assets. The asset pricing literature has seen much empirical evidence establishing the effect of uninformative supply on asset prices. Nazeran (2010b) in particular reviews a number of such papers and provides more evidence form the Treasury bills market.

If the dynamics of an asset market are driven by DSDCs, market participants may not be price-takers anymore. In contrast to a perfectly competitive market, under a DSDC regime, hedge-funds and investment-bankers may have to respond to public information by surreptitious portfolio adjustments. Antitrust courts and bankruptcy courts may have to decree verdicts mindful of the demand-elasticity of the liable party’s assets. Modigliani and Miller (1958) assumption would be violated, affecting
the cost of equity. Also, the cost of a private-equity buyout can be affected by the asset’s demand elasticity.

On another note, non-price-taking market participants may be able to strategize and take excess profit. That could be indicative of market inefficiency, a contentious topic in the literature. To know whether a DSDC leads to market inefficiency or not, we need to devise a theoretical model for the asset market.

This paper will devise a theoretical framework for studying asset prices in a market where assets are perfectly inelastically supplied and elastically demanded. The framework can help with evaluating situations in which market efficiency or inefficiency may arise. The framework can also be used to devise a model, and perhaps calibrate it to the data. The model can then be used to advise investors on the effect of positive or negative supply on an asset’s price.

In section 2, I will look at some of the major asset pricing models and frameworks, notably CAPM, APT, Lucas and SDF. Looking at each model, I seek to answer two questions: First, does the model have an underlying DSDC; second, can the demand curve be conveniently calibrated to that of the market and studied. I will show all, except APT, do have underlying DSDC’s, but only SDF is suitable for studying DSDC’s. In section 3, I propose a theoretical asset pricing framework, suitable for pricing assets in the context of an inelastic supply curve and a downward sloping demand curve. Lastly, the implications, and methods for statically testing a model will be examined.

The proposed framework in this paper can be thought of as a specialized version of the SDF framework; fit to better study a DSDC. This paper’s intention is explaining
and theorizing a seemingly existent downward sloping demand curve in the asset market.

2.2 Literature Review

Before proposing a new framework - in section 3 - I will motivate the need for one here.

Some asset pricing frameworks assume assets are perfectly elastically supplied so demand is irrelevant for pricing. Examples are APT and the Black-Scholes option pricing model. On the other hands, there are frameworks that assume, in the short-run, assets are inelastically supplied. In other words, for any asset $a$, $\sum_{n=1}^{N}q_{n,a,t} = Q_{a,t}$ where $q_{n,a,t}$ is investor $n$’s holding of asset $a$ at time $t$, and $Q_{a,t}$ is the total number of shares of asset $a$. Also $\sum_{a=1}^{A}P_{a,t} \cdot Q_{a,t} = \sum_{n=1}^{N}W_{n,t}$ at time $t$, where $P_{a,t}$ is the price of asset $a$, $W_{n,t}$ is the wealth of investor $n$, $A$ is the number of different assets available, and $N$ is the number of investors. $Q_{a,t}$ is assumed to remain constant for extended durations of time. This paper is interested in this latter type, as they are suitably structured to explain supply effects.

Perhaps the most classical model in this category is CAPM, and then of course the Lucas model, both of which will be looked into before studying the more general SDF framework, as laid out by Hansen and Richard (1987).

2.2.1 CAPM

Analytical Derivation of CAPM

Using Markowitz’s mean-variance efficient portfolio method, we get

$$\omega = \frac{(\bar{R}_p - R_f)V^{-1}(\bar{R} - R_f)e}{(\bar{R} - R_f)e'V^{-1}(\bar{R} - R_f)e} \quad (2.1)$$
where $\omega$ is a vector of weights for each of the $N$ risky assets. $\bar{R}$ is a vector of the expected return of the $N$ assets, and $e$ is a $N$x1 vector of ones. $V$ is the $N$x$N$ variance-covariance matrix for the returns of all assets. $\bar{R}_p$ is expected return of the portfolio along the market capital line, which is chosen by the investor’s utility function. The weight of the riskless asset in the portfolio is $1 - e'\omega$.

For the special case where $e'\omega = 1$, we get the tangency point between the market capital line and the efficient frontier hyperbola, which is called the market portfolio. In that case:

$$\omega = \frac{V^{-1}(\bar{R} - R_f e)}{e'V^{-1}(\bar{R} - R_f e)} \quad (2.2)$$

The covariance of each asset with this market portfolio is a vector:

$$\text{cov}(R,R_m) = V\omega = \frac{(\bar{R} - R_f e)}{e'V^{-1}(\bar{R} - R_f e)} \quad (2.3)$$

The variance of the market portfolio is

$$\text{var}(R_m) = \sigma_m^2 = \omega'V\omega = \frac{\omega'(\bar{R} - R_f e)}{e'V^{-1}(\bar{R} - R_f e)} = \frac{(\bar{R}_m - R_f)}{e'V^{-1}(\bar{R} - R_f e)} \quad (2.4)$$

where $\bar{R}_m$ is the expected return of the market portfolio.

Since the denominators are equal, from the two equations above we conclude the market (tangency) portfolio can be characterized as:

$$\frac{\bar{R}_i - R_f}{\text{cov}(R_i,R_m)} = \frac{\bar{R}_m - R_f}{\text{cov}(R_m,R_m)} \quad (2.5)$$

where $i$ indexes an asset in the portfolio. This result is known as CAPM.
Model Identification

Assume a two asset economy, with both assets in the market portfolio $R_m = wR_i + (1 - w)R_{-i}$. Then both sides of the CAPM equation contain $R_i$.

$$\frac{\bar{R}_i - R_f}{\text{cov}(R_i, R_m)} = \frac{\bar{R}_m - R_f}{\text{cov}(R_m, R_m)} = \frac{w\bar{R}_i + (1 - w)\bar{R}_{-i} - R_f}{\text{var}(wR_i + (1 - w)R_{-i})}$$ (2.6)

bringing $\bar{R}_i$ to the left-hand-side and $\bar{R}_{-i}$ to the right-hand-side, after much simplification we get:

$$\frac{\bar{R}_i - R_f}{w\text{var}(R_i) + (1 - w)\text{cov}(R_i, R_{-i})} = \frac{\bar{R}_{-i} - R_f}{(1 - w)\text{var}(R_{-i}) + w\text{cov}(R_i, R_{-i})}$$ (2.7)

Solving the equation above for $w$, we get Markowitz model’s specification for the tangency portfolio.

$$w = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \cdot V^{-1}(\bar{R} - R_f e)$$ (2.8)

Letting $R_i = \tilde{P}_i$ and $R_{-i} = \tilde{P}_{-i}$ with the distribution for the $\tilde{P}_i$ payoff, the $\tilde{P}_{-i}$ payoff and $R_f$ given, the equation has three unknowns, $P_i$, $P_{-i}$ and $w$. Clearly we don’t get identification. In order to identify the asset prices we need the two market clearing conditions:

$$w \cdot W = Q_i P_i$$ (2.9)

$$(1 - w) \cdot W = Q_{-i} P_{-i}$$ (2.10)

where $W$ is the total market wealth, $Q_i$ is the total number of shares of asset $i$ and $Q_{-i}$ is the total number of shares of asset $-i$ available in the market. Now we have sufficient equations to find the unknowns.
Demand Curve

Models that require market-clearing conditions to price assets usually have underlying demand curves, and it turns out CAPM indeed does have one.

\[
\omega = \frac{(\bar{R}_p - R_f)\mathbf{V}^{-1}(\bar{R} - R_fe)}{(R - R_fe)\mathbf{V}^{-1}(\bar{R} - R_fe)} \tag{2.11}
\]

\[
m \equiv \frac{(\bar{R}_p - R_f)}{(R - R_fe)\mathbf{V}^{-1}(\bar{R} - R_fe)} \tag{2.12}
\]

where \(m\) is a scalar, characterizing the investor’s preference along the market-capital line\(^{25}\). Then in a two asset economy with many investors, an individual holds the following portfolio:

\[
\begin{bmatrix}
w_1 \\
w_2
\end{bmatrix} = m \begin{bmatrix}
\frac{\text{var}(\tilde{P}_1)}{\tilde{P}_1} & \frac{\text{cov}(\tilde{P}_1, \tilde{P}_2)}{\tilde{P}_1} \\
\frac{\text{cov}(\tilde{P}_1, \tilde{P}_2)}{\tilde{P}_1} & \frac{\text{var}(\tilde{P}_2)}{\tilde{P}_2}
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{P}_1 - R_f \\
\tilde{P}_2 - R_f
\end{bmatrix}
\]

\[
= m \begin{bmatrix}
\frac{1}{\tilde{P}_1} & 0 \\
0 & \frac{1}{\tilde{P}_2}
\end{bmatrix}^{-1} \begin{bmatrix}
\text{var}(\tilde{P}_1) & \text{cov}(\tilde{P}_1, \tilde{P}_2) \\
\text{cov}(\tilde{P}_1, \tilde{P}_2) & \text{var}(\tilde{P}_2)
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{P}_1 - P_1R_f \\
\tilde{P}_2 - P_2R_f
\end{bmatrix}
\]

\[
\begin{bmatrix}
w_1 \\
w_2
\end{bmatrix} = m \begin{bmatrix}
\text{var}(\tilde{P}_1) & \text{cov}(\tilde{P}_1, \tilde{P}_2) \\
\text{cov}(\tilde{P}_1, \tilde{P}_2) & \text{var}(\tilde{P}_2)
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{P}_1 - P_1R_f \\
\tilde{P}_2 - P_2R_f
\end{bmatrix}
\]

(2.13)

Knowing \(w_1 \cdot W = q_1 P_1\) where \(W\) is the investor’s wealth and \(q_1\) is the number of shares he owns of asset 1, we see \(\frac{w_1}{\tilde{P}_1} = \frac{q_1}{W}\). So:

\[
mW \begin{bmatrix}
\text{var}(\tilde{P}_1) & \text{cov}(\tilde{P}_1, \tilde{P}_2) \\
\text{cov}(\tilde{P}_1, \tilde{P}_2) & \text{var}(\tilde{P}_2)
\end{bmatrix}^{-1} \begin{bmatrix}
P_1R_f \\
P_2R_f
\end{bmatrix}
\]

\[
= mW \begin{bmatrix}
\text{var}(\tilde{P}_1) & \text{cov}(\tilde{P}_1, \tilde{P}_2) \\
\text{cov}(\tilde{P}_1, \tilde{P}_2) & \text{var}(\tilde{P}_2)
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{P}_1 \\
\tilde{P}_2
\end{bmatrix} - \begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
\]

\[
\Rightarrow R_f \begin{bmatrix}
P_1 \\
P_2
\end{bmatrix} = \begin{bmatrix}
\tilde{P}_1 \\
\tilde{P}_2
\end{bmatrix} - \frac{1}{mW} \begin{bmatrix}
\text{var}(\tilde{P}_1) & \text{cov}(\tilde{P}_1, \tilde{P}_2) \\
\text{cov}(\tilde{P}_1, \tilde{P}_2) & \text{var}(\tilde{P}_2)
\end{bmatrix} \begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
\]

(2.14)

(2.15)

\(^{25}\)In the special case where the net market holding of bonds is zero, the net market holding of risky assets is represented by the tangency portfolio and hence called the market portfolio and \(\frac{1}{m} = e'\mathbf{V}^{-1}(\bar{R} - R_fe)\).
In non-vector notation we have two downward sloping demand curves:

\[ P_1 R_f = \bar{P}_1 - \frac{1}{mW} (q_1 \text{var}(\tilde{P}_1) + q_2 \text{cov}(\tilde{P}_1, \tilde{P}_2)) \]  
\[ P_2 R_f = \bar{P}_2 - \frac{1}{mW} (q_2 \text{var}(\tilde{P}_2) + q_1 \text{cov}(\tilde{P}_1, \tilde{P}_2)) \]  

(2.16)  
(2.17)

**Substitution**

Markowitz optimization comes from the first order condition of an unspecified utility function with a specific structure. The portfolio \( w \) is optimal in the sense that \( \frac{dU}{dw} = 0 \). However this does not implicate that \( \frac{dP}{dq} = 0 \). The demand curve is neither globally nor locally perfectly elastic; however, if \( q_1 P_1 \ll W \), a change in \( q_1 \) will have a negligible effect, so at the margin the demand curve would be *virtually* flat.

To summarize, CAPM models an asset market with DSDC, but asymptotically the demand curve becomes flat. CAPM restricts the slope of the demand curve to the variance of future payoffs, hence limiting our ability to calibrate modeled demand curves to that of the market.

**2.2.2 APT**

Portfolio choice goes hand in hand with sloped demand curves. APT, although a more generalized version of CAPM, doesn’t have portfolio implications. What makes CAPM a portfolio choice model is the definition of its factor, namely, the market portfolio. Factors containing portfolio information can give APT a demand curve as well. For example, Fama and French (1993) three factor model changes equation 5 to:

\[ \bar{R}_i - R_f = \frac{\text{cov}(\bar{R}_i, \bar{R}_m)}{\text{var}(\bar{R}_m)} (\bar{R}_m - R_f) + \lambda_1 f_1 + \lambda_2 f_2 \]  

(2.18)
where \( f_1 \) and \( f_2 \) are the SMB and HML factors, and \( \lambda_1 \) and \( \lambda_2 \) are the factor loadings.

Substituting in the market portfolio, like equation 6, the equivalent of equation 8 becomes:

\[
 w = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \cdot V^{-1}(\bar{R} - R_f e) \\
 e'V^{-1}(\bar{R} - R_f e) \\
 - \frac{\lambda_1 f_1 + \lambda_2 f_2}{1 - w} \cdot \frac{(w^2 \text{var}(R_i) + wcov(R_i, R_{-i}) + (1 - w)\text{var}(R_{-1}))}{e'V^{-1}(\bar{R} - R_f e)} 
\]

Clearly the equation cannot be analytically solved for \( w \), hence the demand curve cannot be analytically derived. The significance of this limitation will be explained in the next section.

All in all, APT, per se, doesn’t have an underlying demand curve, but based on the definition of its factors, it may have one. This doesn’t mean the demand curve will have an analytical representation.

2.2.3 Lucas Model

The model assigns a value to a payoff, using the Inter-temporal Marginal Rate of Substitution (IMRS). The value is a function of future consumption distribution, but not the level of exposure to the asset’s risk. The price does not vary with the quantity of shares bid, and this is expected as the Lucas model is not a portfolio choice model, at least explicitly. Since the price is \( P_t = E[P_{t+1} P_t^{*}] \), where \( P_t^{*} \) is the IMRS; on the face of it the demand curve is perfectly elastic, and so the Lucas model doesn’t serve as a proper platform. This would, however, be a hasty conclusion.

As depicted in Figure 1, a supply shock should not affect equilibrium price. On the other hand, more trees on a Lucas island result in more consumption, smaller \( U'(C_{t+1}) \) and consequently lower prices. In other words, as illustrated in Figure 2, a
Figure 2.1: Perfectly Elastic Demand with Perfectly Inelastic Supply

Figure 2.2: The Effect of Positive Supply Shock on Asset Prices
supply shock from 1 to 2 causes a demand shift from 1 to 2. Consequently, we get a DSDC.

So does the Lucas model have a DSDC? Yes, but it will be shown that it is analytically cumbersome, and sometimes impossible, to theoretically study DSDC’s in the Lucas framework.

In a Crusoe Economy with one asset we have \( v_t = \max\{u(C_t) + \beta \int v_{t+1} f(s', s) ds\} \), subject to \( W_{t+1} = R_{t+1} \cdot (W_t - C_t) \), where \( R_{t+1} = \frac{P_{t+1} + dt_{t+1}}{P_t} \). In the notation of Lucas (1982) the investment amount \( W_t - C_t \) can be written as \( q_t P_t \), where \( q_t \) is the quantity of shares purchased at \( P_t \). Hence, \( W_{t+1} = q_t \cdot (P_{t+1} + dt_{t+1}) \) and \( C_t + q_t P_t \leq W_t \). The Euler equation is

\[
P_t = \beta E\left[ \frac{U'(C_{t+1})}{U'(C_t)} \right] (P_{t+1} + dt_{t+1})
\]

but knowing \( C_t = W_t - q_t P_t \) (for increasing utility functions), \( W_t = q_{t-1} (P_t + dt_t) \), and assuming \( U'(C_t) = C_t^{-\sigma} \) we have:

\[
P_t = \beta E\left[ \frac{C_t^{\sigma}}{C_{t+1}^{\sigma}} (P_{t+1} + dt_{t+1}) \right] = \beta E\left[ \frac{(W_t - q_t P_t)^\sigma}{(W_{t+1} - q_{t+1} P_{t+1})^\sigma} (P_{t+1} + dt_{t+1}) \right]
\]

\[
P_t = \beta E\left( \frac{q_{t-1} dt_t - \Delta q_t P_t}{q_t dt_{t+1} - \Delta q_{t+1} P_{t+1}} \right)^\sigma (P_{t+1} + dt_{t+1})
\]

Equation 22 can neither be solved for \( q_t \), nor can it be solved for \( P_t \). This is the primary difficulty with this model. In order to go any further, we would either need to make the not so suitable assumption that \( \Delta q_{t+1} = 0 \), or make the restrictive assumption that \( \sigma = 1 \). Under the former assumption, we can solve for quantity but not price:

\[
q_t = \frac{W_t}{P_t \left( \frac{P_t}{\beta E\left[ \frac{P_{t+1} + dt_{t+1}}{dt_{t+1}} \right]^\sigma} \right)^{\frac{1}{\sigma}}}
\]

(2.22)
We are facing a specific class of demand curves, notably those of the form. This is because of the power utility form we assumed. But for an arbitrary demand curve of interest, an analytical form of the utility function may not exist or be hard to derive. This would be the second limitation.

Under the \( \sigma = 1 \) assumption, we can solve for price, but not quantity:

\[
P_t = \frac{q_t - 1 \Delta q_t + 1 \beta E[P_{t+1} + d_{t+1}]}{\Delta q_t + \beta E[q_{t+1} + d_{t+1} + \Delta q_{t+1}]} \quad (2.23)
\]

\( \sigma = 1 \) implies log utility, exhibiting another limitation of the model. Moreover, both \( E[P_{t+1} + d_{t+1}] \) and \( E[q_{t+1} + \Delta q_{t+1} P_{t+1}] \) are endogenous (unless in a two-period context) which makes it harder to judge the real shape of the demand curve.

For analytical tractability, throughout the rest of this section, we will assume \( \sigma = 1, \Delta q_{t+1} = 0 \Rightarrow q_t = \frac{W_t}{P_t (1 + \frac{1}{\beta E[|F|]})} \). In presence of a second investor, Mr.Friday, both investors must face the same \( P_t \), therefore,

\[
1 q_t^R = \frac{W_t^R}{1 + \frac{1}{\beta E[|F|]}} \quad (2.24)
\]

Assuming information symmetry, then \( \frac{q_t^R}{q_t^F} = \frac{W_t^R}{W_t^F} \). By market clearing condition,

\[
q_t^R + q_t^F = Q_t \Rightarrow q_t^R = \frac{W_t^R}{1 + \frac{1}{\beta E[|R|]}} \cdot Q_t \quad (2.25)
\]

This equation exhibits Lucas model’s portfolio-choice implication. Note it has an analytical representation only because of the restrictive assumptions made. Assuming Robinson and Friday have heterogeneous conditioning information, then

\[
q_t^R = \frac{W_t^R (1 + \frac{1}{\beta E[|F|]})}{W_t^R (1 + \frac{1}{\beta E[|F|]}) + W_t^F (1 + \frac{1}{\beta E[|R|]})} \cdot Q_t
\]

\[
\Rightarrow P_t = \frac{1}{Q_t} \left( \frac{W_t^R}{1 + \frac{1}{\beta E[|F|]}} + \frac{W_t^F}{1 + \frac{1}{\beta E[|F|]}} \right) \quad (2.26)
\]
Equation 27 shows the Lucas demand curve which can be solved for both quantity and price.

In a two asset economy, with similar assumptions as above:

\[
P_{1,t} = \beta E\left[\frac{U'(C_{t+1})}{U'(C_t)} (P_{1,t+1} + d_{1,t+1})\right] = \beta E\left[\frac{C_t}{C_{t+1}} (P_{1,t+1} + d_{1,t+1})\right]
\]

\[
P_{1,t} = E\left[\frac{W_t - q_{1,t}P_{1,t} - q_{2,t}P_{2,t}}{q_{1,t}(P_{1,t+1} + d_{1,t+1}) + q_{2,t}(P_{2,t+1} + d_{2,t+1}) - q_{1,t+1}P_{1,t+1} - q_{2,t+1}P_{2,t+1}}\right] \cdot \beta
\]

\[
P_{1,t} = \beta E\left[\frac{W_t - q_{1,t}P_{1,t} - q_{2,t}P_{2,t}}{q_{1,t}(P_{1,t+1} + d_{1,t+1}) + q_{2,t}(P_{2,t+1} + d_{2,t+1}) - q_{1,t+1}P_{1,t+1} - q_{2,t+1}P_{2,t+1}}\right]
\]

\[
\Rightarrow P_{1,t} = \frac{W_t - q_{2,t}P_{2,t}}{q_{1,t}d_{1,t+1} + q_{2,t}d_{2,t+1}} \cdot \beta\]

Equation 29 shows how introducing a new asset (tree) \( P_{2,t} \) reduces the price of the first asset, \( P_{1,t} \), unless when \( q_{2,t} \) is negative. An investor will choose to short asset 2 and long asset 1 when shorting asset 2 can hedge asset 1’s risk; where the introduction of asset 2 can increase the value of asset 1. If the investor chooses to long both asset, asset 1’s price will drop. The forth difficulty is lack of parameterization for calibrating the effect of \( q_{2,t}P_{2,t} \) on \( P_{1,t} \).

Also, to note about equation 29 is that in presence of two assets, even with log-utility, it cannot be solved for \( q_{1,t} \). Plugging equation 29 for \( q_{2,t}P_{2,t} \) into equation 29 for \( q_{1,t}P_{1,t} \) gives equation 30\(^\text{26}\).

\[
q_{1,t}P_{1,t} = \frac{1}{\beta E[\frac{P_{2,t+1} + d_{2,t+1}}{q_{2,t}d_{2,t+1} + q_{1,t}d_{1,t+1}}]} + \frac{1}{\beta E[\frac{P_{2,t+1} + d_{2,t+1}}{q_{2,t}d_{2,t+1} + q_{1,t}d_{1,t+1}}]} + \frac{1}{\beta^2 E[\frac{P_{2,t+1} + d_{2,t+1}}{q_{2,t}d_{2,t+1} + q_{1,t}d_{1,t+1}}]}
\]

The demand curve cannot be solved showing the fifth limitation which arises in presence of multiple assets. Unquestionably, in a multiple investor and multiple assets context, an analytical form of the demand curve becomes all the more unreachable.

\(^{26}\)\( q_{2,t} = \alpha_2 t q_{1,t} \) and \( \alpha_{1,t} = \frac{1}{\alpha_{2,t}} \).
How can the Lucas model have portfolio choice implications, and a downward sloping demand curve, while satisfying absence of arbitrage? The model assumes with a wealth of $W_t$, $q_tP_t$ will be invested so that $\sum_t U(W_t - q_tP_t)$ is maximized. $C_t$ is introduced to proxy the more notationally cumbersome $W_t - q_tP_t$. The investor invests what she invests, but our hope is she will consume the rest; therefore, as a matter of convenience we choose the letter $C$. Consumption is also there to make a story for why the maximization object is $U(W_t - q_tP_t)$, which makes the agent risk-averse over the asset’s payoff. It is important to understand consumption isn’t really there and the mathematics of the model is only conscious of wealth and investment. Risk aversion over payoff, in addition to $W_t - q_tP_t > 0$ which puts a bound on $q_t$; overall give us a downward sloping demand curve without arbitrage. The significance of $U(W_t - q_tP_t)$ and a bound on $q_t$ will be discussed more in section 3.1, after theorem 3.1.1.

As a last criticism to this model, I claim institutional investors, and market makers, as the strong forces in the market’s price discovery mechanism, are not driven by their client’s consumption motives but rather the firm’s profit-maximizing objective. Even though consumers can pick a mutual-fund given their risk-aversion, since most mutual-fund investments are long term (or retirement) investments, consumption based models shouldn’t be tested with monthly, quarterly, or perhaps even annual data.
2.2.4 SDF

Chamberlain (1983) proves that under general conditions, asset pricing models must all be of the form \( P_t = E[P_{t+1}P^*_t] \) where \( P^*_t \) is the Stochastic Discount Factor (SDF). Hansen and Richard (1987) (heretoforth HR 87) shows the conditioned version of Chamberlain also holds. This section will follow in HR 87’s footstep. The framework introduced in section 3 overlaps with the SDF framework, and under reasonable assumption, is a subset of SDF asset pricing framework. The SDF framework however, at HR 87’s level of abstraction, does not have a demand curve. I posit a \( q_t \) should be in \( P^*_t \). In subsections 2.4.1 and 2.4.2, I will show the need for \( q_t \) in presence of heterogeneous beliefs, from two different perspectives. Explanation about the theoretical roots of the \( P_t = E[P_{t+1}P^*_t(q_t)] \) form simplifies the understanding of section 3 framework. Subsection 2.4.3 will provide a critique of mainstream interpretation of risk-free returns, which will then be addressed in the new framework.

**Conditioning Sigma-Algebra**

Since in a market with homogeneous investors (in information and preference) no one will take a short position, allowing for heterogeneity of information is an important feature for an asset pricing model. Short positions are an undeniable part of the market transactions and sometimes necessary for achieving equilibrium. Investor heterogeneity of information and preferences will be an important part of the framework introduced in section 3. Here I will show conditions under which the SDF framework can handle heterogeneous investors.

\(^{27}\) Not because anyone claimed \( P^*_t \) is not a function of \( q_t \), but simply because I don’t know of any paper that showed it explicitly.
Let Robinson be an investor with rational expectations, knowing the true measures of each state of the world, with idiosyncratic conditioning information, represented by his sigma-algebra $\Sigma^R$. His price for the $P_{t+1}$ payoff is a random variable measurable in $\Sigma R$ $P_t^R = E[P_{t+1}P_t^*|\Sigma^R]$. Clearly another agent, Mr.Friday, with a different conditioning sigma-algebra $\Sigma^F$, can not observe $P_t^R$. Likewise, Robinson cannot observe $P_t^F$, as it is measurable in $\Sigma^F$. On the other hand, at equilibrium, both investors must agree on the market price, $P_t$. To converge these two seemingly contradicting conditions, let $R_o$ be the smallest element in $\Sigma^R$, which contains the true state, and likewise define $F_o$. Knowing $P_t^R$ and $P_t^F$ are mappings from $\Sigma^R$ and $\Sigma^F$, respectively, to $\mathbb{R}$, for the market price to be agreed by all, we must have $P_t^R(R_o) = P_t^F(F_o)$. But since all agree on the true measures of the states of the world, when $R_o \neq F_o$ they may disagree on the measures of $R_o$ and $F_o$. Hence the equality $E[P_{t+1}P_t^*|R_o] = E[P_{t+1}P_t^*|F_o]$ holds if $P_{t+1}^*$ is a function of a control variable. I posit the risk exposure level as an investor’s control variable, so by the market clearing forces, the market comes to equilibrium price $P_t$. Concretely, at equilibrium $P_t = E[P_{t+1}P_t^*(q_t^R)|\Sigma^R] = E[P_{t+1}P_t^*(q_t^F)|\Sigma^F]$ and $q_t^F + q_t^R = Q_t$.

**Unconditional Testing**

The major contribution of HR 87 is not just proving the conditional version of SDF but also to show it can still be tested by an econometrician with unconditional information. In other words, if $P_t^R = E[P_{t+1}P_t^*|\Sigma^R] \Rightarrow P_t = E[P_t^R] = E[E[P_{t+1}P_t^*|\Sigma^R]] = E[P_{t+1}P_t^*]$ by the law of iterated expectations. This subsection calls the literature’s attention to a seldom noticed caveat, but will precede that with a schematic of HR 87 proof.
Let $Z_t = \{\forall P_{t+1} \text{ s.t. } P_t(P_{t+1}) = 0\}$, and let $P_t(P_{t+1})$ be the pricing function, pricing the payoff $P_{t+1}$ at time $t$. Further assume $P_t(\cdot)$ is a linear function so the price of a portfolio is the portfolio of prices. Take an arbitrary asset $P_o$, and construct $r_{t+1}^* = \frac{P_{o,t+1} - Z(P_{o,t+1})}{P_t(P_{o,t+1})}$, where $Z(P_{o,t+1})$ is the zero priced component of $P_{o,t+1}$ with the least second moment.

$P_{t+1}$ can be represented as an infinite dimensional vector, in a Hilbert space, where each dimension represents a state of the world. A vector’s component along a dimension is the corresponding random-variable’s value in that state of the world. The dot-product will be defined as the conditional-expectation of the product. $Z_t$ would be a linear sub-space, and $Z(\cdot)$ would be the projection operator. In other words $Z(P_{t+1})$ is $P_{t+1}$’s projection on the $Z_t$ space, or equivalently its zero priced component with the least second moment. Evidently $r_{t+1}^*$ is orthogonal to $Z_t$ by construction. Take any $P_{t+1}$ and construct $P_{t+1} - P_t(P_{t+1})r_{t+1}^*$ and we have a vector in the $Z_t$ subspace. Taking its dot-product with the orthogonal vector $r_{t+1}^*$, adopting Dirac’s bra-ket notation, we get:

$$
\langle P_{t+1} - P_t(P_{t+1})r_{t+1}^* | r_{t+1}^* \rangle_{\Sigma^R} = 0 \Rightarrow \langle P_{t+1} | r_{t+1}^* \rangle_{\Sigma^R} = P_t(P_{t+1}) \langle r_{t+1}^* | r_{t+1}^* \rangle_{\Sigma^R}
$$

$$
\Rightarrow P_t(P_{t+1}) = \left( \langle P_{t+1} | \frac{r_{t+1}^*}{\langle r_{t+1}^* | r_{t+1}^* \rangle_{\Sigma^R}} \rangle_{\Sigma^R} \right)_{\Sigma^R} \tag{2.30}
$$

The SDF is defined as $P_{t+1}^* \equiv \frac{r_{t+1}}{E[r_{t+1}^* r_{t+1}^*]_{\Sigma^R}}$ in $P_t(P_{t+1}) = E[P_{t+1} \frac{r_{t+1}}{E[r_{t+1}^* r_{t+1}^*]_{\Sigma^R}}]_{\Sigma^R}$.

The caviot is that taking the unconditional expectation may be challenged by Jensen’s inequality.

The significance is that an econometrician can test a model with unconditional information only if $P_{t+1}^*$ is constructed to be independent of conditioning information, for example, $P_{t+1}^* = \frac{C_t}{C_{t+1}}$. I will show; such definition of $P_{t+1}^*$ will lead to inconsistent
pricing unless $P_{t+1}^*$ is a function of control variables. In order to see why, first a definition, two lemmas and a theorem are in order.

**Definition:** Given vector $x$, let vector $y = \frac{x}{\langle x| x \rangle}$. $y$ is the geometric inverse of $x$, i.e. $y \equiv x^{-1}$.

**Lemma 1:** If $y$ is $x$’s geometric inverse, then the geometric inverse of $y$ is $x$. i.e. $y = x^{-1} \rightarrow x = y^{-1}$.

Proof: $\frac{y}{\langle y| y \rangle} = \frac{x}{\langle x| x \rangle} \frac{\langle x| x \rangle}{\langle x| x \rangle} = x$. QED.

**Lemma 2:** If $y$ is $x$’s geometric inverse, $\langle x| y \rangle = \langle x| x \rangle \langle y| y \rangle = 1$

Proof: $y = \frac{x}{\langle x| x \rangle} \rightarrow \langle y| y \rangle = \frac{\langle x| y \rangle}{\langle x| x \rangle} \rightarrow \langle x| y \rangle = \langle x| x \rangle \langle y| y \rangle = \langle x| x \rangle \left( \frac{x}{\langle x| x \rangle} \right) = \left( \frac{x}{\langle x| x \rangle} \right)^2 = 1$. QED.

**Theorem:** For any $x$ in a Hilbert space, its geometric inverse is uniquely defined.

Proof: By definition, a vector’s geometric inverse is parallel to it, and only differs by a constant of proportionality. If $\exists x, y, z$ such that, $y = \frac{x}{\langle x| x \rangle}$ and $z = \frac{x}{\langle x| x \rangle}$, then $y$ and $z$ are parallel so must be that $z = \alpha y$ for some scalar $\alpha$. If $\alpha = 1$ the theorem is proved. $z = \alpha y = \frac{x}{\langle x| x \rangle} \rightarrow \langle z| x \rangle = \alpha \langle y| x \rangle = \langle x| x \rangle = 1 \Rightarrow$ using Lemma 2, $\alpha = 1$. QED.

Adopting the above terminology, $P_{t+1}^*$ and $r_{t+1}^*$ are each others’ geometric inverse, and uniquely defined. If one is independent of conditioning information, the other one, inevitably, must be dependent on the conditioning information, by definition of geometric inversion.

If $P_{t+1}^*$ is independent of conditioning information, then for Robinson

$$r_{t+1}^* \leftarrow P_{t+1}^* = \frac{P_{t+1}^*}{\langle P_{t+1}^* | P_{t+1}^* \rangle_{\Sigma R}} = \frac{P_{a,t+1} - Z(P_{o,t+1})}{P_t(P_{o,t+1})}$$

$$\Rightarrow Z(P_{a,t+1}) = P_{o,t+1} - P_t(P_{o,t+1}) \left( \frac{P_{t+1}^*}{\langle P_{t+1}^* | P_{t+1}^* \rangle_{\Sigma R}} \right) \quad (2.31)$$

53
By definition \( Z(P_{t+1}) \) is in \( Z_t \); therefore, \( P_t(Z(P_{t+1})) = 0 \). Also, we know \( P_t^F(P_{t+1}) = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F} \) and \( P_t^R(P_{t+1}) = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R} \). On the other hand when Friday prices \( Z(P_{t+1}) \) at zero, using 31:

\[
0 = \langle P_{t+1}^*|Z(P_{t+1})\rangle_{\Sigma^F} = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F} - P_t \frac{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F}}{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R}} \tag{2.32}
\]

\[
\Rightarrow P_t = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F} \frac{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R}}{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F}} \neq \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F} \tag{2.33}
\]

But for Robinson:

\[
0 = \langle P_{t+1}^*|Z(P_{t+1})\rangle_{\Sigma^R} = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R} - P_t \frac{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R}}{\langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F}} \Rightarrow P_t = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R} \tag{2.34}
\]

So if \( P_{t+1}^* \) is independent of conditioning information, pricing \( Z(P_{t+1}) \) and pricing \( P_{t+1} \) will only be consistent for both investors if \( \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^F} = \langle P_{t+1}^*|P_{t+1}\rangle_{\Sigma^R} \). Similar to 2.4.1, necessary for this equality is to have \( P_{t+1} \) depend on a control variable.

Hence we have \( P_t = E[P_{t+1}^*|q_t^R]_{\Sigma^R} = E[P_{t+1}^*|q_t^F]_{\Sigma^F} \) with three unknowns \( \{P_t, q_t^R, q_t^F\} \) and two equations. If the two control variables \( q_t^R \) and \( q_t^F \) are the exposure level, the market clearing condition \( q_t^R + q_t^F = Q_t \) will be the third and necessary (not sufficient) equation for identification.

Note in Lucas model \( P_{t+1}^* = \left( \frac{C_{t+1}}{C_t} \right)^q \) which by section 2.3 is equals \( \frac{(W_t - q_t P_t)^q}{(W_{t+1} - q_{t+1} P_{t+1})^q} \) so a function of the control variable \( q_t \).

**Risk-Free Return**

A fixed-income asset with \( \text{Prob}\{P_{t+1} = 1\} = 1 \) would be priced as \( P_t = E[P_{t+1}^*] \); therefore the risk-free return is \( \frac{1}{E[P_{t+1}^*]} \). More generally however, any asset with an expected payoff of one, and no correlation with \( P_{t+1}^* \) will be priced \( \frac{1}{E[P_{t+1}^*]} \), and there can be many such assets. In fact in the SDF framework, the term “risk-free” refers to a broader class of assets than the fixed-income assets whereas in the Markowitz
model, the risk-free asset is defined to have zero-variance; indeed risk-free. In this case, “risk-free” is a misnomer as it implies fixed-income. Alternative names used are “zero-premium” or “systematic-risk-free”. This paper will adopt the “systematic-risk free” name.

This paper would attempt to define the risk-free rate as the interest-rate which is the yield on a fixed-income (zero-variance) asset; and the price that clears the default-free lending and borrowing market at a given maturity. Unlike the Lucas model - as will be addressed in section 3.2 - the price will be unique at any given time, for a given maturity. Consequently, given an economy and a maturity, there can only be a unique time-series of risk-free rates. This definition better matches the definition of the fixed-income T-bills, which are often used to proxy for the systematic-risk-free assets.

Besides that, I claim, adjusted for default risk, the market prices a fixed-income asset strictly higher than a “systematic-risk free” asset with similar expected, yet uncertain, payoff. One objective of the proposed framework in section 3 is to distinguish between these two assets.

2.3 Proposed Model

Many individual investors invest in the stock-market through pension-funds, mutual funds or 401(k)’s. The pension-funds in return invest their money with hedge-funds. The portfolio managers at these financial institutions are risk-averse profit-maximizers with little, if any, knowledge of their clients’ - or worse, clients’ clients (pension-fund member) - consumption smoothing motives. On the other hand, it is hard to imagine popular examples of direct investment like Warren Buffet or Bill
Gates, as well as less famous ones, taking their IMRS or Transversality condition into consideration when investing.

Investor risk-aversion is an important component of my framework in that it contributes to the simultaneous existence of a downward sloping demand curve and absence of arbitrage. The idea of investor risk-aversion is supported by diversification and the existence of Hedge Funds, and a great market for Credit Default Swaps (CDS).

The goal is to introduce an asset pricing framework in which:

- investment decision is made by a risk-averse profit-maximizing institutional investors
- an asset’s price will be bound by the asset’s supply
- heterogeneous information still lead to an equilibrium (refer to section 2.4.1)
- the risk-free asset be defined as a fixed income asset (refer to section 2.4.3)
- with rate of return that clears the money market.

At the center of the framework is a demand curve with a perfectly inelastic supply curve. Sellers in the stock market form an upward slopping supply curve, which is not this inelastic supply but is supply due to intra-market trading. Since the demand, given a quantity, is one’s value for the item, one can think of this intra-market supply curve as the reflection of the demand-curve across the inelastic market supply.

### 2.3.1 Model Description

An investor’s utility is defined as the area under the demand curve. In other words, the utility from a certain exposure level to a payoff is the most the investor is willing to pay for it. Hence the utility is payoff denominated, with diminishing marginal-utility due to the downward slope of the demand curve so the profit-maximizing agent’s
Figure 2.3: Perfectly Inelastic Supply vs. Intra-market Supply

risk-aversion is reflected in its diminishing demand. For a riskless asset (Dirac-delta distribution) with a certain next period payoff of $P_{\text{riskless}}$, the highest an investor is willing to pay for $q$ shares, is $qP_{\text{riskless}}$ and we will consider that the utility derives from that investment. For a risky asset $\tilde{P}_{\text{risky}}$, with a tilde signifying a random-variable with uncertainty, there will be a premium:

$$U(q\tilde{P}_{\text{risky}}) = q\tilde{P}_{\text{risky}} - h(q\tilde{P}_{\text{risky}})$$  \hspace{1cm} (2.35)

where $h(.)$ is a nonlinear function, and zero-valued when the payoff has Dirac-delta distribution. To this extent, the utility has been restricted to have a linear and nonlinear component, where the nonlinear component is zero for riskless assets.

The investor has a wealth $W$. The cash left is a riskless asset, and has a utility. If $\tilde{P}_{\text{risky}}$ is traded at the price $P_{\text{risky}}$, the utility from the cash left is riskless $U(W - qP_{\text{risky}}) = W - qP_{\text{risky}}$.

Assumption 3.1.1 The utility function is linearly additive in riskless payoffs, where riskless means Dirac-delta distribution (zero-variance).
This implies \( U(q_1 \Priskless + q_2 \Prisky_1 + q_3 \Prisky_2) = U(q_1 \Priskless) + U(q_2 \Prisky_1 + q_3 \Prisky_2) \). Therefore for a risky asset, \( U(q \bar{P} + (W - qP)) = q \bar{P} - h(q \bar{P}) + (W - qP) \).

Knowing there is always cash left, the shorthand notation will be \( U(q \bar{P}) = q \bar{P} - h(q \bar{P}) + (W - qP) \). In presence of \( A \) risky assets, utility is

\[
U(\sum_{a=1}^{A} q_a \bar{P}_a) = \sum_{a=1}^{A} q_a \bar{P}_a - h(\sum_{a=1}^{A} q_a \bar{P}_a) + (W - \sum_{a=1}^{A} q_a P_a) \quad (2.36)
\]

In presence of multiple investors, with lending and borrowing at the risk-free rate \( R_f \), for \( N \) investors we have: \( n = 1..N \)

\[
U^n(\sum_{a=1}^{A} q^n_a \bar{P}_a) = \sum_{a=1}^{A} q^n_a \bar{P}_a - h^n(\sum_{a=1}^{A} q^n_a \bar{P}_a) + (W^n - \sum_{a=1}^{A} q^n_a P_a) \cdot R_f \quad (2.37)
\]

Expected utility maximization implies

\[
\frac{d}{dq^n_j} E[U^n|\Sigma^n] = 0 \Rightarrow P_j R_f = E[\bar{P}_j|\Sigma^n] - i^n_j(\sum_{a=1}^{A} q^n_a \bar{P}_a) \quad (2.38)
\]

where \( \Sigma^n \) is investor \( n \)'s sigma-algebra and \( i^n_j(\cdot) \equiv \frac{d}{dq^n_j} E[h(\cdot)|\Sigma^n] \). By construction \( i(\cdot) \) is the non-constant term of marginal utility and hence what gives rise to a downward slope in the marginal utility (demand) curve, and in that sense, the marginal risk-premium. Evidently, a concave utility would imply and monotone increasing \( i_j(\cdot) \) in \( q_j \).

**Theorem 3.1.1** Assuming the risk-free rate is unaffected by shifting a risky asset’s payoff, absence of arbitrage implies \( i(\cdot) \) is invariant under shifting the payoff.

**Proof:** Let \( b \) index one of the assets. If news comes that owners of this asset \( b \) will receive an extra dollar, in addition to the payoff, once the payoff is realized, absence of arbitrage requires price of \( b \) to increase by \( \frac{1}{R_f} \). Before the news \( P_b R_f = E[\bar{P}_b|\Sigma^n] - i^n_b(\sum_{a=1}^{A} q^n_a \bar{P}_a) \). After the news \( P^*_b R_f = E[\bar{P}^*_b|\Sigma^n] - i^n_b(\sum_{a=1}^{A} q^n_a \bar{P}^*_a) \), where \( P^*_b = P_b + \frac{1}{R_f} \), \( \bar{P}^*_b = \bar{P}_b + 1 \), and \( \bar{P}^*_a = \bar{P}_a \) for all \( a \) except \( a = b \). Plugging in these equations in, we see \( i^n_b(\sum_{a=1}^{A} q^n_a \bar{P}_a) = i^n_b(\sum_{a=1}^{A} q^n_a \bar{P}^*_a) \). QED.

58
Sufficient for this is that \( i(.) \) be a linear function of centered moments of asset payoffs. This is quite suitable as with such an \( i(.) \) a \( P^*_t \) can always be constructed\(^{28}\), where upper bar denotes expected value.

Let \( i(.) \equiv \frac{d}{d\hat{q}_a} var(\sum_a q_a \hat{P}_a) \). Then:

\[
P_t R_f = \hat{P}_t - \alpha \hat{q}_i \hat{P}_i + cov(\hat{P}_i, \sum_{j \neq i} q_j \hat{P}_j) = \hat{P}_t - cov(\hat{P}_i, \beta \sum_j q_j \hat{P}_j)
\]

\[
= \hat{P}_t - E[(\hat{P}_i - \bar{P}_i) \beta \sum_j q_j (\hat{P}_j - \bar{P}_j)]
\]

\[
\kappa \equiv -\alpha \sum_j q_j \hat{P}_j
\]

\[
\hat{P}^* = \frac{\kappa - \hat{\kappa} + 1}{R_f} \Rightarrow E[\hat{P}^*] = \frac{1}{R_f} \cdot cov(\hat{P}_i, \hat{P}^*) = \frac{1}{R_f} (\hat{q}_i var(\hat{P}_i) + cov(\hat{P}_i, \sum_{j \neq i} q_j \hat{P}_j))
\]

\[
\Rightarrow P_t = \hat{P}_t \hat{P}^* + cov(\hat{P}_i, \hat{P}^*) = E[\hat{P}_i \hat{P}^*]
\]

The same logic applies for higher order centered moments, but slightly more complicated. As an example, let’s address the case of a third order centered moment where \( i(.) \equiv -\frac{d}{d\hat{q}_a} skew(\sum_a q_a \hat{P}_a) \). Then:

\[
P_t R_f = \hat{P}_t + \beta \sum_j \sum_k E[(\hat{P}_i - \bar{P}_i) q_j \bar{P}_k (\hat{P}_j - \bar{P}_j)]
\]

\[
= \hat{P}_t + \beta \sum_j \sum_k q_j \bar{P}_k (\hat{P}_j - \bar{P}_j)
\]

\[
\kappa \equiv \beta \sum_j \sum_k q_j \bar{P}_k (\hat{P}_j - \bar{P}_j)
\]

\[
\hat{P}^* = \frac{\kappa - \hat{\kappa} + 1}{R_f} \Rightarrow E[\hat{P}^*] = \frac{1}{R_f} \cdot cov(\hat{P}_i, \hat{P}^*) = \frac{1}{R_f} (\hat{q}_i var(\hat{P}_i) + cov(\hat{P}_i, \sum_{j \neq i} q_j \hat{P}_j))
\]

\[
\Rightarrow P_t = \hat{P}_t \hat{P}^* + cov(\hat{P}_i, \hat{P}^*) = E[\hat{P}_i \hat{P}^*]
\]

Noticeably, the equation above is generalizable to higher orders of centered moments. Moreover, when \( i(.) \) is a linear function of centered moments, since \( cov() \) is a linear function, a \( \hat{P}^* \) can always be constructed.
making the model compatible with the SDF framework. Consequently, the general assumptions made by HR 87 would be held as well. Moreover, risk-premium being a linear function of centered moments has been a long-held assumption in a portion of the empirical finance literature, with intuition rather than theory supporting it. The theorem above, supporting the practice, can lead the path to expanding beyond intuitive empirical investigations.

Note theorem 3.1.1 expresses a necessary condition for absence of arbitrage, not sufficient. Hence, \( i(.) \) being a function of centered moments is a sufficient condition for the necessary condition of absence of arbitrage. In fact the difficulty of constructing an \( i(.) \) consistent with absence of arbitrage may at first skip the eye. The difficulty arises from the existence of the control variable. HR 87 prove \( E[P_{t+1}P^*_t] \) is arbitrage free if and only if \( P^*_t \) is nonnegative at all states of the world. This necessitates a bound on the control variable inside \( P^*_t \), which is just what the Lucas model does. In the Lucas model \( P^*_t \) is a ratio of two positive values (or a monotonic transformation thereof), namely \( C_t \) and \( C_{t+1} \), such that \( 0 \leq C_t \leq W_t - q_t P_t \) (refer to section 2.3). This puts a bound on \( q_t \).

A model in this proposed framework is constituted of a \textit{function form for} \( i(.) \), \textit{its parameterization}, and \textit{bounds on the variables} used in \( i(.) \). The bounds are important for absence of arbitrage.

Adding time subscripts and assuming all investors agree on \( E[P_{j,t+1}\Sigma^n] \) to simplify notation, we get:

\[
P_{j,t}R_{f,t} = E[P_{j,t+1}] - i^n_j(\sum_{a=1}^A q_{a,t}^n P_{a,t+1}) \quad (2.39)
\]

At any time \( t \) there are \( A(N+1)+1 \) unknowns: \( \{P_{a,t}, q_{a,t}^n, R_{f,t}\} \). Equation 39 provides \( A \cdot N \) equations. \( \sum_{n=1}^N q_{a,t}^n = Q_a \), where \( Q_a \) is the total supply of asset \( a \), provides
A more market-clearing conditions. To identify the market it is necessary to also introduce the clearing condition for the lending and borrowing market.

\[
N \sum_{n=1}^{N} (W_n - \sum_{a=1}^{A} q_n^a P_a) = 0 \Rightarrow \\
\sum_{n=1}^{N} W_n = \sum_{n=1}^{N} \sum_{a=1}^{A} q_n^a P_a = \sum_{a=1}^{A} P_a \sum_{n=1}^{N} q_n^a \\
\sum_{n=1}^{N} W_n = \sum_{a=1}^{A} Q_a P_a \tag{2.40}
\]

The formulation above completes the description of the core of the framework. In section 3.3 I will introduce a couple of enhancements, but before that, some implications of the framework are in order.

### 2.3.2 Implications

- For a given asset \( a \), \( E[P_{a,t+1}] - P_{a,t} R_f = i^a_n (\sum_{a=1}^{A} q_{a,t}^n P_{a,t+1}) \). This equation captures the predicted risk-premium equivalence. LHS is agent independent, but the RHS has agent dependent conditioning information and preference parameters in \( i(.) \), as well as control variables \( q_{a,t}^n \). Therefore, \( q_{a,t}^n \) must be chosen so that all agents incur the same risk premium given their conditioning information and preferences. This closely corresponds to the argument in section 2.4.1, as well as conclusion drawn from equation 34 and 35 in section 2.4.2.

- The function form in equation 16 is very similar to the proposed framework which puts asset prices in general form of:

\[
P_n R_f = \bar{P}_n - i\left(\sum_m q_m P_m\right) \tag{2.41}
\]

In case of CAPM:

\[
i\left(\sum_m q_m P_m\right) = \frac{1}{2mW} \frac{d}{dq_n} var\left(\sum_m q_m P_m\right) \tag{2.42}
\]
which for first of two assets becomes
\[
\frac{1}{2mW} \frac{d}{dq_1} \text{var}(q_1 \tilde{P}_1 + q_2 \tilde{P}_2) = \frac{1}{mW}(q_1 \text{var}(\tilde{P}_1) + q_2 \text{cov}(\tilde{P}_1, \tilde{P}_2))
\] (2.43)

Working backwards to the utility function
\[
h(q_1 \tilde{P}_1 + q_2 \tilde{P}_2) \equiv \frac{1}{2mW}((q_1 \tilde{P}_1 + q_2 \tilde{P}_2) - (q_1 \bar{P}_1 + q_2 \bar{P}_2))^2
\] (2.44)
\[
U(q_1 \tilde{P}_1 + q_2 \tilde{P}_2) = q_1 \bar{P}_1 + q_2 \bar{P}_2 - h(.)
\] (2.45)

• By clearing condition for the lending and borrowing market
\[
\sum W^*_n = \sum Q_a P_{a,t} = \sum \frac{Q_a}{R_{f,t}} (E[P_{a,t+1}] - i^a_n(.))
\]
\[
\Rightarrow R_{f,t} = \frac{\sum_{a=1}^A Q_a \cdot (E[P_{a,t+1}] - i^a_n(.))}{\sum_{n=1}^N W^*_n}
\] (2.46)
meaning an increase in aggregate wealth, or the risk-premium \(i\), reduces \(R_f\).

An increase in \(N\) (new investors; perhaps foreign investors, or Fed’s quantitative easing) also reduces \(R_f\). An increase in \(A\) (new investments), \(Q_a\) or \(E[P_{a,t+1}]\) for any \(a\), ceteris paribus, increases \(R_f\).

• \(R_{f,t} = E[R_{a,t+1}] - \frac{i^a_n(\cdot)}{P_{a,t}}\), where \(R_{a,t+1} = \frac{P_{a,t+1}}{P_{a,t}}\), assuming \(i^a_n(\cdot)\) is not a function of \(R_f\), which given its purpose is reasonable. The equation says the price that clears the default-free lending and borrowing market is equal to any asset’s return minus the risk-premium per unit price of that asset. When every agent optimally chooses how many shares of an asset to own (risk-exposure), \(i^a_n(\cdot)\) is adjusted so that the equation above holds for every asset \(a\).

In the Lucas model however:
\[
P_{a,t} = E[P_{a,t+1}]E[P^*_t] + \text{cov}(P_{a,t+1}, P^*_t) \Rightarrow
\]
\[
P_{a,t} R_{f,t} = E[P_{a,t+1}] + \text{cov}(P_{a,t+1}, P^*_t)R_{f,t} \Rightarrow i \equiv -\text{cov}(P_{a,t+1}, P^*_t)R_{f,t}
\] (2.47)
Equation 48 shows the Lucas model is a special case of our proposed framework, where \( i(\cdot) \) is a constant function, but a function of \( R_f \). Hence:

\[
R_{f,t} = E[R_{a,t+1}] + \frac{\text{cov}(P_{a,t+1}, P_{t+1})}{P_{a,t}} R_{f,t}
\]

\[
\Rightarrow R_{f,t} = E[R_{a,t+1}] + \frac{1}{\text{cov}(P_{a,t+1}, P_{t+1})} - 1 E[R_{a,t+1}]
\]

(2.48)

implying that rather than the premium adjustment to only depend on marginal risk, it is proportional to expected return. This helps compare Lucas’s systematic-risk free return with the money-market clearing risk-free return of \( R_{f,t} = E[R_{a,t+1}] - i_{n}(\cdot) P_{a,t} \).

According to footnote 4, the proposed framework can be put in the SDF notation, but according to the same footnote, \( P_{t+1}^{*} \) would be a function of \( R_f \) and hence different from that of equation 49. This is a result of the different interpretations of a risk-free asset, namely risk-free and systematic-risk-free.

- Rewriting last equation as \( E[R_{a,t+1}] = R_{f,t} + \frac{\gamma_{a}(\cdot)}{P_{a,t}} \) gives the APT form. Coincidentally, Hou, Karolyi, and Kho (2006) finds a list of factor that can explain global stock returns, which starts with earnings over price, cash-flow over price and dividend over price.

### 2.3.3 Enhancements

There are two enhancements to be made to the model. One implicit assumption of the model has been that the investor buys all \( q_a \) shares at the constant price \( P_a \). In other words, even though the agent has a DSDC, its demand is marginal in aggregate demand, hence does not shift aggregate demand upward at its intersection with the inelastic supply. On the other hand, we are particularly interested in institutional
investors with market power, as well as the upward sloping supply curve revealed by
Level II quotes. A simple way to remedy this is assuming market price responds a
purchase request and one has to pay the same price for all purchased assets. In that
case, the expected utility is:

\[
EU(\sum a^n q_{a,t} P_{a,t+1}) = E[\sum a^n q_{a,t} P_{a,t+1}] - Eh(\sum a^n q_{a,t} P_{a,t+1})
\]
\[
+ (W^n_t - \sum a^n q_{a,t} P_{a,t} q_{a,t})R_{f,t}
\]
\[
\Rightarrow 0 = E[P_{j,t+1}] - i^n_j (\sum a^n q_{a,t} P_{a,t+1}) - P_{j,t}(q_{j,t})R_{f,t} - q^n_j R_{f,t} \frac{dP_{j,t}(q^n_{j,t})}{dq^n_{j,t}}
\]
\[
\Rightarrow P_{j,t}(q^n_{j,t})R_{f,t} = \frac{E[P_{j,t+1}] - i^n_j (\sum a^n q_{a,t} P_{a,t+1})}{(1 - \frac{dln(P_{j,t}(q^n_{j,t}))}{dln(q^n_{j,t}))}}
\]

(2.49)

Meaning an investor will attempt to discount the price by the demand elasticity of
price. Under this market structure investors with market-power will exercise to win
excess-profit. On the other hand, if the investor can buy each new share at a different
price as it climbs up the supply curve, q shares cost \( \int_0^q P(q) dq \). Therefore:

\[
EU(\sum q^n_{a,t} P_{a,t+1}) = E[\sum q^n_{a,t} P_{a,t+1}] - Eh(\sum q^n_{a,t} P_{a,t+1})
\]
\[
+ (W^n_t - \sum q^n_{a,t} P_{a,t} q^n_{a,t})R_{f,t}
\]
\[
\Rightarrow 0 = E[P_{j,t+1}] - i^n_j (\sum a q^n_{a,t} P_{a,t+1}) - P_{j,t}(Q_a + q^n_{j,t})R_{f,t}
\]
\[
\Rightarrow P_{j,t} R_{f,t} = E[P_{j,t+1}] - i^n_j (\sum a q^n_{a,t} P_{a,t+1})
\]

This assumption is closer to how the clearing mechanism works at electronic stock
markets, and shows how price discrimination can allow for DSDC’s without exercise
of market power. The new equilibirum price will be that of a competitive market
with price-taking investors.
The second enhancement that can be made is to $q$. The model as laid out above assumes the investor enters each period with no exposure to an asset’s risk. If we assume the investor starts current period with an endowment $q_e$, in the basic setup there is no change:

$$EU\left[\sum q_t P_{t+1}\right] = E\left[\sum q_t P_{t+1}\right] - Eh\left(\sum q_t P_{t+1}\right) + (W_t - \sum (q_t - q_e)P_t)R_{f,t}$$  \hspace{1cm} (2.50)

$$\Rightarrow 0 = E[P_{t+1}] - i(\cdot) - P_tR_{f,t}$$

But if we use the pricing system in equation 50,

$$EU\left[\sum q_t P_{t+1}\right] = E\left[\sum q_t P_{t+1}\right] - Eh\left(\sum q_t P_{t+1}\right) + (W - \sum (q_t - q_e)P_t)R_{f,t}$$

$$\Rightarrow 0 = E[P_{t+1}] - i(\cdot) - P(q_t)R_{f,t} - (q_t - q_e)R_{f,t}\frac{dP_t}{dq_t}$$  \hspace{1cm} (2.51)

$$\Rightarrow P(q_t)R_{f,t}(1 + (1 - \frac{q_e}{q_t})\frac{q_t}{P_t(q_t)}\frac{dP_t}{dq_t}) = E[P_{t+1}] - i(\cdot)$$

$$\Rightarrow P_tR_{f,t} = \frac{E[P_{t+1}] - i(\cdot)}{1 + (1 - \frac{q_e}{q_t})\frac{dln(P_t)}{dln(q_t)}}$$  \hspace{1cm} (2.52)

When $q = q^e$ the investor neither buys, nor sells and the denominator is one, with market prices like one under perfect competition. Otherwise market price is discounted when $q > q^e$ and inflated when $q < q^e$.

Lastly, equation 39 assumed expected utility maximization while there is no reason to be bound by that. Non-expected maximization is a venue one might find rewarding$^{29}$.

### 2.3.4 Empirics

In this section I will layout different empirical methods, without exhausting all possibilities. For notational simplicity I will stick to the baseline framework. This

$^{29}$Refer to Rabin (2000).
paper does not intend to introduce a framework that simplifies statistical testing. Coincidentally it happens to make it harder.

**Testing a Model**

The framework consists of three sets of unknowns $P_a, q^n_a$ and $R_f$. A model hypothesizes a particular structure for $i(.)$. $P_a$ and $R_f$ are market observables, but the framework is less convenient than conventional frameworks since an econometrician needs the portfolio time-series of an investor as well. By construction, $i(.)$ depends on the investor’s [conditioning] information, which is even harder to obtain than the portfolio data. A way around this is to also hypothesize that the investor’s belief about, say $\text{var}(P_{t+1})$, has been the variance of the payoff time-series plus a white-noise process. Now there are enough components to test the hypothesized structure, although a rejection would be a rejection of the joint hypothesis.

**Testing for Risk Sharing**

When panel-data of investor portfolio is available, not only can the above method be applied, but also the model’s hypothesized risk-premium equivalence, joint with the hypothesized structure of $i(.)$, can be tested.

The next couple of tests do not required investor portfolios:

**Supply Shock**

$E[P_{t+1}] - i(\sum q_a P_{a,t+1})$ is [a linear transformation of] an investor’s inverse demand-curve. Solving for $q$, $q = \hat{i}^{-1}(EP_{t+1} - P_tR_f)$ is an investor’s demand curve. Let $q^* = \sum_{n=1}^N q^n = I^{-1}(EP_{t+1} - P_tR_f)$ be the aggregate demand curve. Intersecting the aggregate inverse demand curve $P_t = E[P_{t+1}] - I(q^*)$ with the inelastic supply curve gives $P_t = E[P_{t+1}] - I(Q)$. When there are new shares offered in the market,
or a major investor dumps a substantial portion of its shares in the market, it can be approximated as a change in $Q$. By checking the predicted price change against the actual price change, at the event, the structure of $i(.)$ can be tested.

**Intra-market Supply Curve**

The aggregate demand curve can be fit against the supply and demand curve revealed by Level II quotes. Unfortunately I am not aware of any data set that has historical Level II data. Such database would be orders of magnitude larger than TAQ.

### 2.4 Conclusion

There are empirical evidences for a downward sloping demand curve in the asset market, but its implications are unclear. Theoretically studying such a market is quite restrained, tedious and sometimes impossible with mainstream asset pricing models like CAPM, APT or the Lucas model, but the SDF framework stands capable to explain asset pricing models with a downward sloping demand curve. I develop a framework, one step less abstract that the SDF, with the intention to pave way to an asset pricing model that has an underlying downward sloping demand curve. The framework is also designed to model heterogeneous, risk-averse, institutional investors in an asset market which is side by side of a money-market.

This framework is to solve or tame some of the difficulties of working in the Lucas framework. One criticism however, is it is static, in contrast to Lucas’s dynamic model. But this framework is intended to model institutions, and traditionally firm’s are modeled as being intertemporally risk-neutral. Besides, Lucas model becomes
dynamic becomes of consumers’ consumption motive, which doesn’t apply to institutions. One inter-temporal compromise investors do need to make however is the payment of dividend to shareholders. An institutional investor always faces the option to making dividend payments, versus keeping the money and investing more. Dividend payments tend to gain more shareholders, and also improve the credit quality as designated by rating institutions, which in return reduces the cost of leveraging for the investor. Modeling dividend payment can be a venue of expanding this framework into a dynamic one.
In an attempt to uncover latent variables, a part of the econometric literature has focused on factor analysis. The most recent developments in the methodology are due to Stock and Watson (2002) and Bernanke, Boivin, and Eliasz (2005) and referred to as “Factor Augmented Vector Auto-Regression” (FAVAR). For asymptotic consistency, FAVAR requires a “large” number of time-series which contain the latent-factor. On other hand, due to the classical-error-in-variable problem, FAVAR is susceptible to small-sample bias. In this paper I develop an unbiased variant of FAVAR which imposes less stringent requirements upon the time-series and can potentially work with as few as 3 time-series. I further show that Principal Components Analysis’ superiority emerges when one has time-series in excess of this minimum.

3.1 Introduction

Many time-series used by econometricians are proxies for the true factors that generate economic data. For example, there is a true inflation factor which affects macroeconomic observables like price or nominal interest rate. On the other hand, what an econometrician observes is a number of proxies like the Bureau of Labor Statistic’s (BLS) Consumer Price Index (CPI), The Bureau of Economic Analysis’
(BEA) GDP Price Deflator, or Dallas Fed’s Trimmed Mean Personal Consumption Expenditure.

Econometrically the true factor is drawn from a probability distribution $z$, while the $i$’th proxy is an observation $\tilde{z}^i = z + u^i$ where $u^i$ is the idiosyncratic disturbance stemming from institutional measurement errors, or sheer unobservability. The random variable $z$ is the latent factor which is to be approximated using the observable $\tilde{z}^i$. Econometricians tend to have proxies of choice that they incorporate into their analysis.

On the other hand, there are hundreds of economic time-series potentially containing informative content, yet in the interest of the degrees-of-freedom, econometricians tend to use less than a dozen at a time. All the while, we have numerous methods at our disposal that can be implemented in order to use all the information content of these proxies, and in turn improve estimation efficiency.

One such method is the Principal Components Analysis (PCA). PCA is an econometric method that orthogonalizes the variance-covariance matrix of a set of time-series and finds the linear-combination where much of the volatility lies. In order to illustrate this method’s ability to clean up the error in proxies, I have generated Figure 3.1. To construct the figure, I use PCA to generate the proxy $\tilde{z}^{PCA}$, and introduce a standard-deviation ratio statistic. The statistic is the ratio of standard deviation of the proxy’s deviation from the latent factor, and the latent factor’s standard deviation. In other words the vertical axis reports $\frac{\sigma(\tilde{z} - z)}{\sigma(z)}$.

Specifically, I start by drawing $T$ observations from a standard normal distribution and assign them to $z$. Then I draw 3 independent sets of $T$ such observations and
Figure 3.1: Ratio of Standard Deviation Statistic

In first column from left, $T = 100$. In second column, $T = 1000$, and in third column, $T = 10,000$. Horizontal axis is $p$, which determines relative volatility of the latent-factor and the idiosyncratic disturbance. In top row, $\tilde{z}_1$ is used as proxy. In middle row, $\tilde{z}_{PC}(1,2)$ is used as proxy. In bottom row, $\tilde{z}_{PC}(1,2,3)$ is used as proxy. The three lines are average value, over 10,000 independent draws, two standard-errors above and two-standard-errors below.
assign them to $u^1$, $u^2$, and $u^3$ respectively. I then construct:

\[ \tilde{z}^1 = pz + u^1 \]  
\[ \tilde{z}^2 = pz + u^2 \]  
\[ \tilde{z}^3 = pz + u^3 \]  
\[ \tilde{z}_{PC}^{(1,2)} = PC(\tilde{z}^1, \tilde{z}^2) \]  
\[ \tilde{z}_{PC}^{(1,2,3)} = PC(\tilde{z}^1, \tilde{z}^2, \tilde{z}^3) \]

The parameter $p$ is inserted to regulate the relative standard deviation of $z$ and each of the noises, and is assigned to the horizontal axis of each graph. In Figure 3.1, $p$ takes on 9 values, from 1 to 5. The three graphs in the left column are made with $T = 100$ observations, those in the middle column are made with $T = 1000$ observations and those to the right are made with $T = 10,000$ observations. The top row shows mean of $\frac{\sigma(z - \tilde{z}^1)}{\sigma(z)}$, with two standard-errors above and two standard-errors below$^{30}$. The middle row shows, $\frac{\sigma(z - \tilde{z}_{PC}^{(1,2)})}{\sigma(z)}$, and the bottom row shows $\frac{\sigma(z - \tilde{z}_{PC}^{(1,2,3)})}{\sigma(z)}$, as a function of $p$.

The nine graphs in Figure 3.1 clearly show that for any number of observations, and any given $p$, PCA improves the proxy’s approximation of the true factor. Further, the bottom row, in comparison to the middle row, shows that inputting 3 proxies to PCA obtains even better results. But more observations (moving from left columns to right) only reduces the standard error of the standard-deviation ratio estimator. Lastly, as $p$ grows, the noise is less relatively volatile, and all three approximators of $z$ perform better. Nonetheless, even at $p = 5$, a PCA over three proxies exhibits half the standard-deviation ratio of when $\tilde{z}^1$ is used as the proxy.

$^{30}$The mean and standard errors are calculated from 10,000 Monte-Carlo simulations
Some proxies are constructed to directly approximate an unobservable, like the growth rate of BLS’s survey of CPI which proxies inflation. Other latent variables however may be hidden in other observables. For instance, real-productivity growth may be hidden in nominal interest rate which is also believed to contain information about inflation. From an econometric stand point, the situation can be illustrated with three observable random variables \( y^1, y^2, \) and \( y^3 \) whose data-generating-processes (DGP) are:

\[
\begin{align*}
    y^1 &= w^1 \beta^1 + p^1 z + u^1 \\
    y^2 &= w^2 \beta^2 + p^2 z + u^2 \\
    y^3 &= w^3 \beta^3 + p^3 z + u^3
\end{align*}
\] (3.6) (3.7) (3.8)

Extracting a PCA-enhanced estimation of \( z \) out of the three equations above is slightly more involved, in that it is a two step process. In step one, if the three \( w^i \)'s can be assumed observable and uncorrelated with \( z, u^1, u^2, \) and \( u^3 \), by simply regressing the three \( y^i \)'s on the three \( w^i \)'s one can get an estimate of the \( \tilde{z}^i \)'s above as:

\[
\tilde{z}^i = y^i - w^i \hat{\beta}^i = \hat{p}^i z + \hat{u}^i
\] (3.9)

In this paper, any reference to an “exogenous regressor” implies both uncorrelated with the disturbance and uncorrelated with the latent factor. An example of exogenous regressor here, is \( w^i \). In step two, the aforementioned procedure can be executed to get a PCA enhanced approximation for \( z \). It is noteworthy however that the PCA approximation will be that of \( z \), up to a constant of proportionality, due to the \( p^i \) coefficients. If \( p^1 = p^2 = p^3 = p \), the constant of proportionality will trivially be \( \frac{1}{p} \), since the approximated factor is \( pz \).
When the coefficients of $z$ are not all equal, the econometrician will approximate $p^{PCA}z$, where $p^{PCA}$ is not identifiable. Curious about how the approximated $p^{PCA}z$ affects the three $y^i$’s, one may be interested in estimating $\frac{p}{p^{PCA}}$. This requires regressing $y^i$ on $z^{PCA}$, which is not as trivial as it may initially seem.

Elaborating on the challenge of using PCA-generated regressors, and prescribing a solution, is this paper’s objective. The rest of the paper is laid out as following. Section II reviews the literature focusing on some of the major and recent contribution in this literature. In this section I also administer a horse-race between the predominant factor-based method in the literature, and the one proposed in this paper. Section III continues the discussion above by explaining my proposal for how PCA-enhanced time-series should be incorporated in regressions, whether the dependent-variable is auto-regressive or not. Section IV formulates this paper’s take on Vector-Auto-Regression (VAR), specifying the procedure used to facilitate the aforementioned horse-race. I then conclude the paper with a summary.

### 3.2 Literature Review

Perhaps the most highly cited paper in this literature is Stock and Watson (2002). The paper characterizes the “dynamic factor model” by two equations.

\[
X_t = \Lambda F_t + \epsilon_t \tag{3.10}
\]

\[
y_t = \beta_F F_t + \beta_w w_t + \epsilon_t \tag{3.11}
\]

$x_t$ is an N-dimensional vector of time-series and $\epsilon_t$ is a $N \times 1$ vector of idiosyncratic disturbances. $w_t$ is an m-dimensional vector, and $F_t$ is a k-dimensional vector. The $m$ observable factors are components of the $w_t$ vector and the $k$ latent factors are components of the $F_t$ vector. Stock and Watson’s major contribution is relaxing the
iid assumption for the error terms, $e_t$, allowing both cross-sectional and temporal correlations in error terms. This was in contrast to papers like Ding and Hwang (1999) and Belviso and Milani (2006). The Stock and Watson projection procedure involves finding the principal component of the $N$ time-series in $X_t$, and using the $k$ components with highest eigen-value as estimators of $\hat{F}_t$. In the next step, they would find least-square estimates for $\hat{\beta}_F$ and $\hat{\beta}_w$ and use them to project future values of $y_t$.

Since $\hat{F}_t$ would contain some components of $e_t$, $e_t$ would have to be uncorrelated with $\epsilon_t$, to prevent endogeneity. It follows that $e_t$ should not contain $y_t$. Further the $N$ time-series in $X_t$ that contain information about the latent factors should not be auto-regressive, as otherwise the principal-component would be a linear combination of the factors and their lags. At least in the field of Economics, finding non-auto-regressive time-series, that contain the latent factor, but do not contain the time-series of interest, is very restrictive. Further Stock and Watson prove asymptotic consistency when $N \to \infty$, meaning not only do we need to find such time-series, we need many of them. I will show in the next section that a problem with PCA, and proxies in general, is the classical error-in-variable problem and the resulting bias. In this case however, as $N \to \infty$ the magnitude of individual components of a normalized linear-combination becomes so small that the error terms become negligible and hence asymptotic consistency is achieved.

Another highly cited paper in this literature is Bernanke, Boivin, and Eliasz (2005). The paper builds on that of Stock and Watson with a slight improvement. The econometric model is specified as below:

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t$$
\[
\begin{bmatrix}
F_t \\
Y_t
\end{bmatrix} = \Phi (L) \begin{bmatrix}
F_{t-1} \\
Y_{t-1}
\end{bmatrix} + v_t
\]

\(X_t\) is a vector of \(N\) time-series, \(F_t\) is a vector of \(K\) latent factors and \(Y_t\) is a vector of \(M\) observables of interest. This paper improves on the earlier one by addressing the earlier objection regarding \(e_t\) having to be void of \(y_t\). The authors calculate the principal components of \(X_t\), then regress them on \(Y_t\) and define the residual as the new set of principal-components. Except for the mentioned improvement, this specification is still subject to the earlier stated criticisms.

Bernanke et al. propose two methods, one based on PCA and one based on Bayesian which is more reliant on the VAR structure. In many cases, the latent factor may be the aggregation of many true macroeconomic factors, virtually all of which could be auto-regressive.\(^{31}\) Hence, as much as \(F_t\) is expected to be auto-correlated, it is difficult to believe it would be a finite order auto-regressive process. For this reason the success of the FAVAR is sensitive to the presence of all building-blocks of \(F_t\) in \(Y_t\), as otherwise an omitted variable problem could lead to biased estimation. Later on, in the VAR section, I will define \(F_t\) specifically to avoid imposing an AR(p) structure upon an otherwise almost certainly auto-correlated latent process. Rather I allow \(F_t\) to be a linear combination of both processes included in \(Y_t\) and those excluded, and instead aggregate all the latent factors in \(z_t\), upon which no structure has been imposed.

While Bernanke et al. describe \(N\) as being “large”, in particular “\(K + M << N\)”, Bai and Ng (2006) prove that under their set of assumptions, factor-augmented regressions “are \(\sqrt{T}\) consistent and asymptotically normal if \(\frac{\sqrt{T}}{N} \to 0\)”. As most

\(^{31}\)If \(A_t\) and \(B_t\) are defined as \((1 - aL)A_t = v_t\) and \((1 - bL)B_t = w_t\), where \(v_t\) and \(w_t\) are iid and white-noise, there does not necessarily exist a \(c\) such that \((1 - cL)(A_t + B_t) = u_t\) where \(u_t\) is also white-noise.
time-series at least go back to the 70’s, a 33 year time-series of monthly data would have $\sqrt{T}$ of almost 20. What $N$ would be large enough for the distributions to be feasibly asymptotic? Moench (2008) chooses 160 time-series, where in his case $\sqrt{T}$ is less than 16. Is 1/10th close enough to 0?

Finding such large number of $N$’s is not easy. Even if it was, Boivin and Ng (2006) show that sometimes an $N$ as low as 40 performs better than one as large as 147. They posit two explanations. First is that cross-correlation in idiosyncratic errors reduces the effectiveness of a large set of latent-factor-containing time-series. Second is that a factor with forecasting power maybe present in a small set of these time-series, and be dominated when $N$ grows too large.

In Section III, I will show that when errors are not cross-correlated, one can be used as instrument, and the other as regressor, to make an unbiased estimate of the factor loadings, not even needing a “large” number of them. Boivin and Ng show when they are correlated, larger $N$’s can be counter-productive. Therefore, I developed an alternative method of augmented latent-factor analysis, which can be executed with smaller $N$.

### 3.2.1 Horse Race

Before getting into details, to showcase the method, I will run a horse-race between the proposed method and the two major approaches in the literature, namely those by Stock and Watson and Bernanke et al.. To conduct the race, I have constructed an arbitrary structure which is to be assumed as the true structure. It is then independently simulated 10,000 times, estimated each time. The mean estimation, bias, standard-error and root-mean-square (RMS) deviation from truth are
then extracted, and reported. Also, each method’s ability to forecast future results, as well as its ability to draw impulse-response-functions (IRF), is put to test. The true structure is as follows:

\[
W_t^i = 0.7 \cdot W_{t-1}^i + 2 \cdot w_t^i, \\
i = 1, 2, 3
\]

\[
Z_t = 0.9 \cdot Z_{t-1} + z_t
\]

\[
F_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} W_{t-1}^1 \\ W_{t-1}^2 \\ W_{t-1}^3 \end{bmatrix} + Z_t
\]

\[
\begin{bmatrix} Y_t^1 \\ Y_t^2 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.4 \\ -0.3 & 0.4 \end{bmatrix} \begin{bmatrix} Y_{t-1}^1 \\ Y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} -0.3 & 0.4 \\ 0.6 & -0.5 \end{bmatrix} \begin{bmatrix} Y_{t-2}^1 \\ Y_{t-2}^2 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} W_t^1 \\ W_t^2 \\ W_t^3 \end{bmatrix}
\]

\[
+ \begin{bmatrix} 3 & 2 & 1 \\ 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} W_{t-1}^1 \\ W_{t-1}^2 \\ W_{t-1}^3 \end{bmatrix} + \begin{bmatrix} 0.3 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} F_{t-1}^1 \\ F_{t-1}^2 \end{bmatrix} + \begin{bmatrix} y_{t-1}^1 \\ y_{t-1}^2 \end{bmatrix}
\]

\[
X_t^i = F_t + 5 \cdot x_t^i, \\
i = 1 \cdots 80
\]

\[
\dot{X}_t^i = F_t + Y_t^1 + 5 \cdot x_t^i, \\
i = 1 \cdots 20
\]

All lowercase variables are independent standard normally distributed and they are the time-series that were generated in each simulation iteration. \(Y^1\) and \(Y^2\) are the variables of interest, \(F\) is the factor, the three \(W^i\)'s are the three exogenous regressors, the \(X^i\)'s are the “slow-moving” factor containing time-series, and \(\dot{X}^i\) are the “fast-moving” ones. The quoted terminology was introduced by Bernanke et al. and was to point out the lagged or contemporaneous response to a shock to \(Y^1\).
Both major methods in the literature first approximate $\hat{F}_t$ using the 100 $X^i$'s and then estimate the equations for $Y^1$ and $Y^2$. The methods differ in how they approximate $\hat{F}_t$ yet both suffer from a classical-error-in-variable problem because of how $\hat{F}_t$ is approximated. Moreover, it is noteworthy that $F_t$ depends on observables $Y^1$ and $Y^2$ and itself is only unobservable because the coefficients of $Y^1$ and $Y^2$ are unobserved and because $Z_t$ is unobservable. This paper’s proposal takes these two observations into consideration.

The proposed method fundamentally hinges on 1) using instruments; 2) thinking of $Z_t$ as the latent factor, rather than $F_t$. Since instruments can be applied while still regarding $F_t$ as the latent factor, I will first run a horse-race between the Stock and Watson method, and an instrumentized version of it, where just the first feature of the proposed model is applied. In the race, I also include the Bernanke et al. method and an instrumentized version of it. A description of the four algorithms is as follows:

- The first implemented method is that of Stock and Watson:
  - Run a PCA on the $T \times 100$ matrix $[X^i_t \hat{X}^i_t]$, and assign it as $\hat{F}_t$.
  - Regress $Y^1_t$ on $[Y^1_{t-1} Y^2_{t-1}, [W^1_t W^2_t W^3_t], [Y^1_{t-2} Y^2_{t-2}], [W^1_{t-1} W^2_{t-1} W^3_{t-1}]$ and $\hat{F}_t$.
  - Regress $Y^2_t$ on $[Y^1_{t-1} Y^2_{t-1}], [W^1_t W^2_t W^3_t], [Y^1_{t-2} Y^2_{t-2}], [W^1_{t-1} W^2_{t-1} W^3_{t-1}]$ and $\hat{F}_{t-1}$.

- Second implementation is an instrumentized version of Stock and Watson:
  - Run a PCA on half the time-series in $[X^i_t \hat{X}^i_t]$, and assign it as $\hat{F}_t$.
  - Run a PCA on the other half of the time-series in $[X^i_t \hat{X}^i_t]$, and assign it as $\hat{I}_t$.  
79
- Regress \( Y_t^1 \) on \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{F}_t \), using \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{I}_t \) as instruments.

- Regress \( Y_t^2 \) on \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{F}_{t-1} \), using \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{I}_{t-1} \) as instruments.

- Third method is due to Bernanke et al.:
  - Run a PCA on the 80 “slow-moving” set, and assign it as \( \hat{F}^*_t \).
  - Run a PCA on the \( T \times 100 \) matrix \([X_t^i \hat{X}_t^i]\), and assign it as \( \hat{F}_t \).
  - Regress \( \hat{F}_t \) on \( \hat{F}^*_t \) and \( Y_t^1 \), and assign \( \hat{F}_t - \beta Y_t^1 \) to \( \hat{F}_t \).
  - Regress \( Y_t^1 \) on \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{F}_t \).
  - Regress \( Y_t^2 \) on \([Y_{t-1}^1 Y_{t-2}^2], [W_t^1 W_{t-1}^2 W_{t-1}^3], [Y_{t-1}^1 Y_{t-2}^2], [W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) and \( \hat{F}_{t-1} \).

- Fourth implementation is the instrumentalized version of Bernanke et al. method:
  - Pick half the “slow-moving” time-series, run a PCA and assign to \( \hat{F}^*_t \).
  - Pick half the “fast-moving” time-series, aggregate with those from the previous item, and assign their PC to \( \hat{F}_t \).
  - Regress \( \hat{F}_t \) on \( \hat{F}^*_t \) and \( Y_t^1 \), and assign \( \hat{F}_t - \beta Y_t^1 \) to \( \hat{F}_t \).
  - Do same as above three items to the other half of the time-series and assign final result to \( \hat{I}_t \).
- Regress $Y^1_t$ on $[Y^1_{t-1} \ Y^2_{t-1} \ W^1_t \ W^2_t \ W^3_t]$, $[Y^1_{t-2} \ Y^2_{t-2}]$, $[W^1_{t-1} \ W^2_{t-1} \ W^3_{t-1}]$ and $\hat{F}_t$, using $[Y^1_{t-1} \ Y^2_{t-1}]$, $[W^1_t \ W^2_t \ W^3_t]$, $[Y^1_{t-2} \ Y^2_{t-2}]$, $[W^1_{t-1} \ W^2_{t-1} \ W^3_{t-1}]$ and $\hat{I}_t$ as instruments.

- Regress $Y^2_t$ on $[Y^1_{t-1} \ Y^2_{t-1} \ W^1_t \ W^2_t \ W^3_t]$, $[Y^1_{t-2} \ Y^2_{t-2}]$, $[W^1_{t-1} \ W^2_{t-1} \ W^3_{t-1}]$ and $\hat{F}_{t-1}$, using $[Y^1_{t-1} \ Y^2_{t-1}]$, $[W^1_t \ W^2_t \ W^3_t]$, $[Y^1_{t-2} \ Y^2_{t-2}]$, $[W^1_{t-1} \ W^2_{t-1} \ W^3_{t-1}]$ and $\hat{I}_{t-1}$ as instruments.

The result for this race is reported in Table 3.1 and Table 3.2. The first table reports results for $Y^1$ and the latter table reports results for $Y^2$.

If one were to plugin the equation for $F$, into the equations for $Y^1$ and $Y^2$, we would get an alternative specification of the same structure:

$$
\begin{bmatrix}
Y^1_t \\
Y^2_t
\end{bmatrix} =
\begin{bmatrix}
1.1 & -0.1 \\
-0.3 & 0.4
\end{bmatrix}
\begin{bmatrix}
Y^1_{t-1} \\
Y^2_{t-1}
\end{bmatrix} +
\begin{bmatrix}
-0.3 & 0.4 \\
0.8 & -0.3
\end{bmatrix}
\begin{bmatrix}
Y^1_{t-2} \\
Y^2_{t-2}
\end{bmatrix}
$$

$$
+ 
\begin{bmatrix}
2 & 1.6 & 3.9 \\
1 & 2 & 3
\end{bmatrix}
\begin{bmatrix}
W^1_t \\
W^2_t \\
W^3_t
\end{bmatrix} +
\begin{bmatrix}
3 & 2 & 1 \\
1 & 3.4 & 2.6
\end{bmatrix}
\begin{bmatrix}
W^1_{t-1} \\
W^2_{t-1} \\
W^3_{t-1}
\end{bmatrix}
$$

$$
+ 
\begin{bmatrix}
0.3 & 0 \\
0 & 0.2
\end{bmatrix}
\begin{bmatrix}
Z^1_t \\
Z^2_{t-1}
\end{bmatrix} +
\begin{bmatrix}
y^1_t \\
y^2_t
\end{bmatrix}
$$

Estimating this alternative specification is more involved. Both a regressor and an instrument for $\hat{Z}_t$ must be approximated. The steps taken in the proposed method are as follows:

- Proposed method:

  - Regress each “slow-moving” $X^i_t$ on $[Y^1_{t-1} \ Y^2_{t-1}]$ and $[W^2_t \ W^3_t]$ using $[W^2_{t-1} \ W^3_{t-1}]$ and $[W^2_t \ W^3_t]$ as instruments, and assign the residual to $\hat{Z}_t$. 

81
– Regress each “fast-moving” \( X_i t \) on \( Y_{t-1}^1, Y_{t-2}^1 \), and \([W_t^2 W_t^3]\) using \( W_t^1 \), \([W_{t-1}^2 W_{t-1}^3]\), and \([W_t^2 W_t^3]\) as instruments,\(^{32}\) and assign the residual to \( \hat{Z}_t \).

– For each “slow-moving” \( X_i t \):
  * Take one of the other 99 \([\hat{Z}_t \hat{Z}_d]\) as instrument \( \hat{I}_t \) and take the PC of the rest and assign it to the regressor \( \hat{R}_t \).
  * Regress each “slow-moving” \( X_i t \) on \( \hat{R}_t, Y_{t-1}^1 Y_{t-1}^2 \), and \([W_t^2 W_t^3]\) using \( \hat{I}_t, Y_{t-1}^1 Y_{t-1}^2 \) and \([W_t^2 W_t^3]\) as instruments, and assign \( X_i t - [Y_{t-1}^1 Y_{t-1}^2 W_t^2 W_t^3] \beta \) to \( \hat{Z}_t \).
  * Regress each “fast-moving” \( X_i t \) on \( \hat{R}_t, Y_t^1, Y_{t-1}^1 Y_{t-1}^2 \), and \([W_t^2 W_t^3]\) using \( \hat{I}_t, Y_t^1, Y_{t-1}^1 Y_{t-1}^2 \) and \([W_t^2 W_t^3]\) as instruments, and assign \( X_i t - [Y_{t-1}^1 Y_{t-1}^2 W_t^2 W_t^3] \beta \) to \( \hat{Z}_t \).

– Repeat step above once more.

– Take half the \([\hat{Z}_t \hat{Z}_d]\) and assign their PC to \( \hat{R}_t \).

– Take the other half and assign its PC to \( \hat{I}_t \).

– Regress \( Y_t^1 \) on \([Y_{t-1}^1 Y_{t-1}^2]\), \([W_t^1 W_t^2 W_t^3]\) and \( \hat{R}_t \), using \([Y_{t-1}^1 Y_{t-1}^2]\), \([W_t^1 W_t^2 W_t^3]\) and \( \hat{I}_t \) as instruments.

– Regress \( Y_t^2 \) on \([Y_{t-1}^1 Y_{t-1}^2]\), \([W_t^1 W_t^2 W_t^3]\), \( \hat{R}_{t-1}, Y_{t-2}^1 Y_{t-2}^2 \), and \([W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) using \([Y_{t-1}^1 Y_{t-1}^2]\), \([W_t^1 W_t^2 W_t^3]\), \( \hat{I}_{t-1}, Y_{t-2}^1 Y_{t-2}^2 \), and \([W_{t-1}^1 W_{t-1}^2 W_{t-1}^3]\) as instruments.

\(^{32}\)Note in this DGP, as apposed to the previous one, the coefficient of \( W_1^1 \) was left zero in the \( F_t \)'s equation. That was to leave \( W^1 \) available as an instrument for \( Y_t^1 \). Separate simulations not reported here show that leaving that coefficient to 1, but instead only relying on the “slow-moving” time-series still yields unbiased results, with negligibly less efficient standard-errors.
The proposed method as well as that of Stock and Watson’s method and Bernanke et al.’s method, under the alternative specification, are compared in Table 3.3 and Table 3.4.

Finally the five different methods are compared in their ability to forecast, as well as their ability to chart impulse-response-functions.

As the first four tables show, using instruments helps correct the bias. This is of course of no surprise as instruments are known to correct the error-in-variable problem. As the forecast-race shows, making the effort to approximate \( \hat{Z}_t \) improves forecasting power. Forecasting is perhaps the area where the proposed method is most useful. Had \( Z_t \) and \( F_t \) been observable, at the \( T + 1 \) horizon, the RMS would have been 1. While Bernanke et al.’s method achieves 1.226, with a roughly 35% improvement the proposed method achieves 1.073.

Lastly, the impulse-response-functions for all five methods meet the qualitative underpinnings of the structure. If an econometrician is solely interested in the qualitative features of the IRF’s, neither method stands out as particularly superior.

Before I discuss the technical details that lead to bias correction in the proposed method, it should be noted that bias correction is not without cost. In other words the proposed method does not strictly dominate the Bernanke et al. method. Although in the structure above, using the proposed method both corrects the bias and reduces standard-error, alternative structures can experience loss of efficiency. The focus of this paper is bias correction but it must be noted that if an econometrician has RMS in mind as the criterion of choice, the proposed method may underperform Bernanke et al.. The implications of the proposed method on standard-error are beyond the
Table 3.1: Bias in $Y_t$’s Regression Coefficients Generated Using Stock and Watson Method, its Instrumentized Variation, Bernanke et al. Method, and its Instrumentized Variation

<table>
<thead>
<tr>
<th>Method</th>
<th>Variant</th>
<th>$Y^1_{t-1}$</th>
<th>$Y^2_{t-1}$</th>
<th>$Y^1_{t-2}$</th>
<th>$Y^2_{t-2}$</th>
<th>$W^1_t$</th>
<th>$W^2_t$</th>
<th>$W^3_t$</th>
<th>$W^1_{t-1}$</th>
<th>$W^2_{t-1}$</th>
<th>$W^3_{t-1}$</th>
<th>$F_t$</th>
<th>$F_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-0.107</td>
<td>0.000</td>
<td>0.028</td>
<td>-0.035</td>
<td>-0.178</td>
<td>-0.159</td>
<td>-0.371</td>
<td>-0.264</td>
<td>-0.174</td>
<td>-0.084</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Orig.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6932</td>
<td>0.3997</td>
<td>0.2725</td>
<td>0.3648</td>
<td>1.8217</td>
<td>0.8413</td>
<td>2.6286</td>
<td>2.7361</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0245]</td>
<td>[0.0186]</td>
<td>[0.0062]</td>
<td>[0.0046]</td>
<td>[0.0233]</td>
<td>[0.0486]</td>
<td>[0.0766]</td>
<td>[0.0332]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1095]</td>
<td>[0.0186]</td>
<td>[0.0282]</td>
<td>[0.0355]</td>
<td>[0.1798]</td>
<td>[0.1660]</td>
<td>[0.3792]</td>
<td>[0.2660]</td>
</tr>
<tr>
<td></td>
<td>S. and W.</td>
<td>-0.126</td>
<td>-0.014</td>
<td>0.029</td>
<td>-0.037</td>
<td>-0.186</td>
<td>-0.193</td>
<td>-0.430</td>
<td>-0.275</td>
<td>-0.181</td>
<td>-0.086</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6744</td>
<td>0.4136</td>
<td>0.2709</td>
<td>0.3634</td>
<td>1.8136</td>
<td>0.8066</td>
<td>2.5705</td>
<td>2.7253</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0277]</td>
<td>[0.0209]</td>
<td>[0.0063]</td>
<td>[0.0048]</td>
<td>[0.0243]</td>
<td>[0.0542]</td>
<td>[0.0864]</td>
<td>[0.0344]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1286]</td>
<td>[0.0250]</td>
<td>[0.0298]</td>
<td>[0.0369]</td>
<td>[0.1880]</td>
<td>[0.2008]</td>
<td>[0.4381]</td>
<td>[0.2768]</td>
</tr>
<tr>
<td></td>
<td>Instr.</td>
<td>0.015</td>
<td>0.015</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.029</td>
<td>0.045</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.029</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8150</td>
<td>0.3851</td>
<td>0.3003</td>
<td>0.3999</td>
<td>2.0001</td>
<td>1.0294</td>
<td>3.0446</td>
<td>2.9091</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0217]</td>
<td>[0.0219]</td>
<td>[0.0064]</td>
<td>[0.0046]</td>
<td>[0.0228]</td>
<td>[0.0468]</td>
<td>[0.0655]</td>
<td>[0.0332]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0264]</td>
<td>[0.0265]</td>
<td>[0.0064]</td>
<td>[0.0046]</td>
<td>[0.0228]</td>
<td>[0.0553]</td>
<td>[0.0793]</td>
<td>[0.0332]</td>
</tr>
<tr>
<td></td>
<td>B., B. and E.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8000</td>
<td>0.4001</td>
<td>0.3000</td>
<td>0.4001</td>
<td>2.0002</td>
<td>0.9999</td>
<td>3.0004</td>
<td>3.0006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0235]</td>
<td>[0.0237]</td>
<td>[0.0065]</td>
<td>[0.0047]</td>
<td>[0.0229]</td>
<td>[0.0501]</td>
<td>[0.0708]</td>
<td>[0.0336]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0235]</td>
<td>[0.0237]</td>
<td>[0.0065]</td>
<td>[0.0047]</td>
<td>[0.0229]</td>
<td>[0.0501]</td>
<td>[0.0708]</td>
<td>[0.0336]</td>
</tr>
</tbody>
</table>

The larger number is the bias. The second number is the average coefficient estimated in the 10,000 Monte-Carlo simulations. The third number (inside square-brackets) is the standard-error of the estimator, according to the simulated results. The forth number (inside curly-brackets) is the root-mean-square deviation from truth.
Table 3.2: Bias in $Y_t^2$’s Regression Coefficients Generated Using Stock and Watson Method, its Instrumentized Variation, Bernanke et al. Method, and its Instrumentized Variation

<table>
<thead>
<tr>
<th>Method</th>
<th>Variant</th>
<th>$Y_{t-1}^1$</th>
<th>$Y_{t-1}^2$</th>
<th>$Y_{t-2}^1$</th>
<th>$Y_{t-2}^2$</th>
<th>$W_{t-1}^1$</th>
<th>$W_{t-1}^2$</th>
<th>$W_{t-1}^3$</th>
<th>$W_{t-1}^4$</th>
<th>$W_{t-1}^5$</th>
<th>$F_t$</th>
<th>$F_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock and Watson</td>
<td>Orig.</td>
<td>-0.055</td>
<td>0.000</td>
<td>0.010</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.020</td>
<td>0.030</td>
<td>-0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3548</td>
<td>0.4000</td>
<td>0.5999</td>
<td>0.5001</td>
<td>1.0002</td>
<td>2.0005</td>
<td>3.0001</td>
<td>0.9995</td>
<td>3.0198</td>
<td>2.0298</td>
<td>0.1898</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0083]</td>
<td>[0.0063]</td>
<td>[0.0208]</td>
<td>[0.0205]</td>
<td>[0.0231]</td>
<td>[0.0229]</td>
<td>[0.0227]</td>
<td>[0.0336]</td>
<td>[0.0536]</td>
<td>[0.0725]</td>
<td>[0.0203]</td>
</tr>
<tr>
<td></td>
<td>Instr.</td>
<td>-0.058</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3580</td>
<td>0.5999</td>
<td>0.5001</td>
<td>1.0002</td>
<td>2.0005</td>
<td>3.0001</td>
<td>1.0006</td>
<td>3.0002</td>
<td>2.0002</td>
<td>0.2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0087]</td>
<td>[0.0064]</td>
<td>[0.0225]</td>
<td>[0.0233]</td>
<td>[0.0230]</td>
<td>[0.0229]</td>
<td>[0.0337]</td>
<td>[0.0563]</td>
<td>[0.0770]</td>
<td>[0.0221]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>{0.0554}</td>
<td>{0.0063}</td>
<td>{0.0231}</td>
<td>{0.0229}</td>
<td>{0.0231}</td>
<td>{0.0227}</td>
<td>{0.0336}</td>
<td>{0.0572}</td>
<td>{0.0784}</td>
<td>{0.0227}</td>
<td></td>
</tr>
<tr>
<td>Bernanke et al.</td>
<td>Orig.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.019</td>
<td>0.029</td>
<td>-0.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2999</td>
<td>0.5999</td>
<td>0.5001</td>
<td>1.0002</td>
<td>2.0005</td>
<td>3.0001</td>
<td>0.9996</td>
<td>3.0189</td>
<td>2.0285</td>
<td>0.1814</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0054]</td>
<td>[0.0064]</td>
<td>[0.0209]</td>
<td>[0.0206]</td>
<td>[0.0232]</td>
<td>[0.0229]</td>
<td>[0.0228]</td>
<td>[0.0336]</td>
<td>[0.0538]</td>
<td>[0.0727]</td>
<td>[0.0272]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{0.0054}</td>
<td>{0.0064}</td>
<td>{0.0230}</td>
<td>{0.0227}</td>
<td>{0.0230}</td>
<td>{0.0229}</td>
<td>{0.0336}</td>
<td>{0.0570}</td>
<td>{0.0781}</td>
<td>{0.0329}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instr.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3000</td>
<td>0.5999</td>
<td>0.5001</td>
<td>1.0002</td>
<td>2.0005</td>
<td>3.0001</td>
<td>1.0006</td>
<td>3.0002</td>
<td>2.0002</td>
<td>0.2002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0054]</td>
<td>[0.0064]</td>
<td>[0.0225]</td>
<td>[0.0233]</td>
<td>[0.0230]</td>
<td>[0.0229]</td>
<td>[0.0337]</td>
<td>[0.0563]</td>
<td>[0.0770]</td>
<td>[0.0221]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>{0.0054}</td>
<td>{0.0064}</td>
<td>{0.0225}</td>
<td>{0.0233}</td>
<td>{0.0230}</td>
<td>{0.0229}</td>
<td>{0.0338}</td>
<td>{0.0563}</td>
<td>{0.0770}</td>
<td>{0.0221}</td>
<td></td>
</tr>
</tbody>
</table>

The larger number is the bias. The second number is the average coefficient estimated in the 10,000 Monte-Carlo simulations. The third number (inside square-brackets) is the standard-error of the estimator, according to the simulated results. The forth number (inside curly-brackets) is the root-mean-square deviation from truth.
Table 3.3: Bias in $Y_t^1$’s Regression Coefficients Generated Using Stock and Watson Method, Bernanke et al. Method, and the Proposed Method.

<table>
<thead>
<tr>
<th>Method</th>
<th>$Y_{t-1}^1$</th>
<th>$Y_{t-1}^2$</th>
<th>$Y_{t-2}^1$</th>
<th>$Y_{t-2}^2$</th>
<th>$W_{t}^1$</th>
<th>$W_{t}^2$</th>
<th>$W_{t}^3$</th>
<th>$W_{t-1}^1$</th>
<th>$W_{t-1}^2$</th>
<th>$W_{t-1}^3$</th>
<th>$Z_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock and Watson</td>
<td>-0.012</td>
<td>0.010</td>
<td>0.028</td>
<td>-0.035</td>
<td>0.173</td>
<td>0.145</td>
<td>0.076</td>
<td>-0.264</td>
<td>-0.174</td>
<td>-0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0876</td>
<td>0.9000</td>
<td>0.2725</td>
<td>0.3648</td>
<td>2.1726</td>
<td>1.7446</td>
<td>3.9761</td>
<td>2.7361</td>
<td>1.8257</td>
<td>0.9159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0052]</td>
<td>[0.0065]</td>
<td>[0.0062]</td>
<td>[0.0046]</td>
<td>[0.0339]</td>
<td>[0.0354]</td>
<td>[0.0343]</td>
<td>[0.0332]</td>
<td>[0.0352]</td>
<td>[0.0405]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{0.0135}</td>
<td>{0.0119}</td>
<td>{0.0282}</td>
<td>{0.0355}</td>
<td>{0.1759}</td>
<td>{0.1488}</td>
<td>{0.0835}</td>
<td>{0.2660}</td>
<td>{0.1778}</td>
<td>{0.0933}</td>
<td></td>
</tr>
<tr>
<td>Bernanke et al.</td>
<td>-0.013</td>
<td>-0.014</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.028</td>
<td>-0.042</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0870</td>
<td>0.1143</td>
<td>0.3003</td>
<td>0.3999</td>
<td>1.9990</td>
<td>1.5724</td>
<td>3.8585</td>
<td>2.9991</td>
<td>1.9988</td>
<td>0.9982</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0202]</td>
<td>[0.0205]</td>
<td>[0.0064]</td>
<td>[0.0046]</td>
<td>[0.0300]</td>
<td>[0.0491]</td>
<td>[0.0658]</td>
<td>[0.0332]</td>
<td>[0.0377]</td>
<td>[0.0443]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{0.0240}</td>
<td>{0.0250}</td>
<td>{0.0064}</td>
<td>{0.0046}</td>
<td>{0.0301}</td>
<td>{0.0563}</td>
<td>{0.0778}</td>
<td>{0.0332}</td>
<td>{0.0377}</td>
<td>{0.0443}</td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>1.1000</td>
<td>0.1003</td>
<td>0.2999</td>
<td>0.4001</td>
<td>0.0000</td>
<td>1.5999</td>
<td>3.9002</td>
<td>3.0004</td>
<td>2.0003</td>
<td>1.0002</td>
<td>0.2972</td>
</tr>
<tr>
<td></td>
<td>[0.0070]</td>
<td>[0.0100]</td>
<td>[0.0065]</td>
<td>[0.0047]</td>
<td>[0.0229]</td>
<td>[0.0297]</td>
<td>[0.0300]</td>
<td>[0.0336]</td>
<td>[0.0382]</td>
<td>[0.0449]</td>
<td>[0.0221]</td>
</tr>
<tr>
<td></td>
<td>{0.0070}</td>
<td>{0.0100}</td>
<td>{0.0065}</td>
<td>{0.0047}</td>
<td>{0.0229}</td>
<td>{0.0297}</td>
<td>{0.0300}</td>
<td>{0.0336}</td>
<td>{0.0382}</td>
<td>{0.0449}</td>
<td>{0.0223}</td>
</tr>
</tbody>
</table>

The larger number is the bias. The second number is the average coefficient estimated in the 10,000 Monte-Carlo simulations. The third number (inside square-brackets) is the standard-error of the estimator, according to the simulated results. The forth number (inside curly-brackets) is the root-mean-square deviation from truth.
Table 3.4: Bias in $Y_t^2$’s Regression Coefficients Generated Using Stock and Watson Method, Bernanke et al. Method, and the Proposed Method.

<table>
<thead>
<tr>
<th>Method</th>
<th>$Y_{t-1}^1$</th>
<th>$Y_{t-1}^2$</th>
<th>$Y_{t-2}^1$</th>
<th>$Y_{t-2}^2$</th>
<th>$W_{t}^1$</th>
<th>$W_{t}^2$</th>
<th>$W_{t}^3$</th>
<th>$W_{t-1}^1$</th>
<th>$W_{t-1}^2$</th>
<th>$W_{t-1}^3$</th>
<th>$Z_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. and W.</td>
<td>-0.055</td>
<td>0.000</td>
<td>0.053</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.216</td>
<td>0.177</td>
<td>0.260</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3548</td>
<td>0.4000</td>
<td>0.8531</td>
<td>0.2990</td>
<td>1.0002</td>
<td>2.0004</td>
<td>3.0001</td>
<td>1.2158</td>
<td>3.5765</td>
<td>2.8604</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>-0.009</td>
<td>-0.009</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.018</td>
<td>-0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>-0.009</td>
<td>-0.009</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.018</td>
<td>-0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>-0.009</td>
<td>-0.009</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.018</td>
<td>-0.028</td>
<td></td>
</tr>
<tr>
<td>B., B. and E.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3001</td>
<td>0.3998</td>
<td>0.8002</td>
<td>0.3001</td>
<td>1.0002</td>
<td>2.0005</td>
<td>3.0001</td>
<td>1.0006</td>
<td>3.4007</td>
<td>2.6009</td>
<td>0.1983</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

The larger number is the bias. The second number is the average coefficient estimated in the 10,000 Monte-Carlo simulations. The third number (inside square-brackets) is the standard-error of the estimator, according to the simulated results. The forth number (inside curly-brackets) is the root-mean-square deviation from truth.
Mean and root-mean-square of deviation of the forecast from truth. In order to focus forecasting
test on the factors and variables of interest, the exogenous-regressors are assumed given at the
forecasted horizons.

<table>
<thead>
<tr>
<th>Projected</th>
<th>Statistic</th>
<th>Stock and Watson</th>
<th>Instrumentized</th>
<th>Bernanke et al.</th>
<th>Instrumentized</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y₁ᵀ+1</td>
<td>Mean</td>
<td>0.015</td>
<td>0.015</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>1.517</td>
<td>1.540</td>
<td>1.226</td>
<td>1.226</td>
<td>1.073</td>
</tr>
<tr>
<td>Y₂ᵀ+1</td>
<td>Mean</td>
<td>0.055</td>
<td>0.055</td>
<td>0.041</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.520</td>
<td>2.559</td>
<td>2.083</td>
<td>2.085</td>
<td>1.713</td>
</tr>
<tr>
<td>Y₁ᵀ+2</td>
<td>Mean</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>1.186</td>
<td>1.189</td>
<td>1.164</td>
<td>1.163</td>
<td>1.152</td>
</tr>
<tr>
<td>Y₂ᵀ+2</td>
<td>Mean</td>
<td>0.054</td>
<td>0.054</td>
<td>0.051</td>
<td>0.051</td>
<td>-0.049</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>3.127</td>
<td>3.162</td>
<td>2.702</td>
<td>2.697</td>
<td>2.192</td>
</tr>
<tr>
<td>Y₁ᵀ+3</td>
<td>Mean</td>
<td>0.049</td>
<td>0.049</td>
<td>0.053</td>
<td>0.053</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>3.464</td>
<td>3.490</td>
<td>3.125</td>
<td>3.117</td>
<td>2.524</td>
</tr>
<tr>
<td>Y₂ᵀ+3</td>
<td>Mean</td>
<td>-0.013</td>
<td>-0.012</td>
<td>-0.019</td>
<td>-0.019</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>1.484</td>
<td>1.497</td>
<td>1.362</td>
<td>1.364</td>
<td>1.254</td>
</tr>
<tr>
<td>Y₁ᵀ+4</td>
<td>Mean</td>
<td>0.014</td>
<td>0.014</td>
<td>0.006</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.108</td>
<td>2.131</td>
<td>1.896</td>
<td>1.901</td>
<td>1.636</td>
</tr>
<tr>
<td>Y₂ᵀ+4</td>
<td>Mean</td>
<td>0.045</td>
<td>0.045</td>
<td>0.054</td>
<td>0.054</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>3.747</td>
<td>3.765</td>
<td>3.494</td>
<td>3.486</td>
<td>2.803</td>
</tr>
<tr>
<td>Y₁ᵀ+5</td>
<td>Mean</td>
<td>0.036</td>
<td>0.035</td>
<td>0.032</td>
<td>0.032</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.507</td>
<td>2.527</td>
<td>2.298</td>
<td>2.298</td>
<td>1.943</td>
</tr>
<tr>
<td>Y₂ᵀ+5</td>
<td>Mean</td>
<td>0.026</td>
<td>0.025</td>
<td>0.030</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.698</td>
<td>2.710</td>
<td>2.527</td>
<td>2.521</td>
<td>2.157</td>
</tr>
<tr>
<td>Y₁ᵀ+6</td>
<td>Mean</td>
<td>0.060</td>
<td>0.061</td>
<td>0.037</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>4.627</td>
<td>4.649</td>
<td>4.376</td>
<td>4.375</td>
<td>3.483</td>
</tr>
<tr>
<td>Y₂ᵀ+6</td>
<td>Mean</td>
<td>0.045</td>
<td>0.045</td>
<td>0.049</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.814</td>
<td>2.822</td>
<td>2.670</td>
<td>2.664</td>
<td>2.293</td>
</tr>
<tr>
<td>Y₁ᵀ+7</td>
<td>Mean</td>
<td>0.052</td>
<td>0.053</td>
<td>0.022</td>
<td>0.022</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>5.137</td>
<td>5.163</td>
<td>4.841</td>
<td>4.840</td>
<td>3.878</td>
</tr>
<tr>
<td>Y₂ᵀ+7</td>
<td>Mean</td>
<td>0.034</td>
<td>0.034</td>
<td>0.031</td>
<td>0.030</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>2.943</td>
<td>2.952</td>
<td>2.824</td>
<td>2.821</td>
<td>2.413</td>
</tr>
<tr>
<td>Y₁ᵀ+8</td>
<td>Mean</td>
<td>0.034</td>
<td>0.035</td>
<td>0.018</td>
<td>0.018</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>5.551</td>
<td>5.577</td>
<td>5.250</td>
<td>5.246</td>
<td>4.256</td>
</tr>
<tr>
<td>Y₂ᵀ+8</td>
<td>Mean</td>
<td>0.027</td>
<td>0.028</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>3.205</td>
<td>3.219</td>
<td>3.057</td>
<td>3.057</td>
<td>2.587</td>
</tr>
<tr>
<td>Y₁ᵀ+9</td>
<td>Mean</td>
<td>0.026</td>
<td>0.026</td>
<td>0.033</td>
<td>0.034</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>5.887</td>
<td>5.909</td>
<td>5.609</td>
<td>5.603</td>
<td>4.601</td>
</tr>
<tr>
<td>Y₂ᵀ+9</td>
<td>Mean</td>
<td>0.042</td>
<td>0.043</td>
<td>0.022</td>
<td>0.021</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>3.511</td>
<td>3.529</td>
<td>3.313</td>
<td>3.314</td>
<td>2.795</td>
</tr>
</tbody>
</table>

88
Figure 3.2: Impulse Response Function for a Shock to $Y^1$: True IRF

Top graphs shows the effect of the impulse on $Y^1$. The middle graphs shows the effect on $Y^2$ and bottom graphs shows the effect on $F$, the latent-factor.
Figure 3.3: Impulse Response Function for a Shock to $Y_1$: Stock and Watson Method

Top graphs shows the effect of the impulse on $Y_1$. The middle graphs shows the effect on $Y_2$ and bottom graphs shows the effect on $F$, the latent-factor. Solid dark-blue line is the true IRF. The solid light-green line is the mean, and the two dash-dot light-green lines around indicate the 90% confidence interval.
Figure 3.4: Impulse Response Function for a Shock to $Y^1$. Instrumentalized Stock and Watson Method

Top graphs show the effect of the impulse on $Y^1$. The middle graphs show the effect on $Y^2$, and bottom graphs show the effect on $F$, the latent factor. Solid dark-blue line is the true IRF. The solid light-green line is the mean, and the two dash-dot light-green lines around indicate the 90% confidence interval.
Figure 3.5: Impulse Response Function for a Shock to $Y^1$: Bernanke et al. Method

Top graphs shows the effect of the impulse on $Y^1$. The middle graphs shows the effect on $Y^2$ and bottom graphs shows the effect on $F$, the latent-factor. Solid dark-blue line is the true IRF. The solid light-green line is the mean, and the two dash-dot light-green lines around is indicate the 90% confidence interval.
Figure 3.6: Impulse Response Function for a Shock to $Y^1$: Instrumentized Bernanke et al. Method
Top graphs shows the effect of the impulse on $Y^1$. The middle graphs shows the effect on $Y^2$ and bottom graphs shows the effect on $F$, the latent-factor. Solid dark-blue line is the true IRF. The solid light-green line is the mean, and the two dash-dot light-green lines around is indicate the 90% confidence interval.
Figure 3.7: Impulse Response Function for a Shock to $Y^1$. Proposed Method

Top graphs shows the effect of the impulse on $Y^1$, and the middle graphs shows the effect on $Y^2$. Solid dark-blue line is the true IRF. The solid light-green line is the mean, and the two dash-dot light-green lines around is indicate the 90% confidence interval.
Table 3.6: RMS of deviation of IRF from Truth.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Variable</th>
<th>Stock and Watson</th>
<th>Instrumentized</th>
<th>Bernanke et al.</th>
<th>Instrumentized</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y¹</td>
<td>0.0791</td>
<td>0.0655</td>
<td>0.0338</td>
<td>0.0399</td>
<td>0.0395</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Y¹</td>
<td>0.0978</td>
<td>0.0832</td>
<td>0.0379</td>
<td>0.0445</td>
<td>0.0441</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0316</td>
<td>0.0396</td>
<td>0.0115</td>
<td>0.0132</td>
<td>0.0131</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.1904</td>
<td>0.2097</td>
<td>0.0346</td>
<td>0.0409</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Y¹</td>
<td>0.0727</td>
<td>0.0600</td>
<td>0.0331</td>
<td>0.0389</td>
<td>0.0399</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0498</td>
<td>0.0462</td>
<td>0.0126</td>
<td>0.0147</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.1633</td>
<td>0.1749</td>
<td>0.0300</td>
<td>0.0352</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Y¹</td>
<td>0.0280</td>
<td>0.0224</td>
<td>0.0206</td>
<td>0.0240</td>
<td>0.0267</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0643</td>
<td>0.0530</td>
<td>0.0286</td>
<td>0.0336</td>
<td>0.0336</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.1392</td>
<td>0.1605</td>
<td>0.0451</td>
<td>0.0533</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Y¹</td>
<td>0.0142</td>
<td>0.0156</td>
<td>0.0147</td>
<td>0.0169</td>
<td>0.0180</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0463</td>
<td>0.0359</td>
<td>0.0286</td>
<td>0.0336</td>
<td>0.0340</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.0864</td>
<td>0.1083</td>
<td>0.0479</td>
<td>0.0561</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Y¹</td>
<td>0.0109</td>
<td>0.0188</td>
<td>0.0190</td>
<td>0.0190</td>
<td>0.0190</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0540</td>
<td>0.0468</td>
<td>0.0195</td>
<td>0.0228</td>
<td>0.0227</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.1009</td>
<td>0.1178</td>
<td>0.0365</td>
<td>0.0429</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Y¹</td>
<td>0.0386</td>
<td>0.0314</td>
<td>0.0197</td>
<td>0.0231</td>
<td>0.0239</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0312</td>
<td>0.0272</td>
<td>0.0119</td>
<td>0.0137</td>
<td>0.0137</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.1039</td>
<td>0.1182</td>
<td>0.0293</td>
<td>0.0343</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Y¹</td>
<td>0.0391</td>
<td>0.0262</td>
<td>0.0178</td>
<td>0.0208</td>
<td>0.0211</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0129</td>
<td>0.0106</td>
<td>0.0089</td>
<td>0.0102</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.0950</td>
<td>0.01080</td>
<td>0.0261</td>
<td>0.0303</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Y¹</td>
<td>0.0232</td>
<td>0.0188</td>
<td>0.0137</td>
<td>0.0158</td>
<td>0.0164</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0107</td>
<td>0.0086</td>
<td>0.0079</td>
<td>0.0092</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.0690</td>
<td>0.0796</td>
<td>0.0225</td>
<td>0.0259</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Y¹</td>
<td>0.0142</td>
<td>0.0117</td>
<td>0.0100</td>
<td>0.0113</td>
<td>0.00124</td>
</tr>
<tr>
<td></td>
<td>Y²</td>
<td>0.0079</td>
<td>0.0065</td>
<td>0.0058</td>
<td>0.0066</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>Factor</td>
<td>0.0477</td>
<td>0.0545</td>
<td>0.0164</td>
<td>0.0185</td>
<td></td>
</tr>
</tbody>
</table>

Root-mean-square of deviation of the IRF from truth at different horizons. Results are reported for the two observable variables of interest as well as the factor. The proposed method does not approximate the factor.
3.3 Using PCA Approximations in Regressions

Mostly, an econometrician’s motivation in enhancing its proxies is obtaining improved regression results. Usability of such enhanced proxies as regressors is fundamental to the relevance of any enhancing procedure. Proxies enhanced by PCA however are not immediately useful in regressions. To understand why, I will give a brief description of PCA’s underlying mechanism.

PCA defines $z_{PCA} = [z^1, z^2] \cdot V^1$ where $V^1$ is the eigen-vector of $z^1$ and $z^2$'s variance-covariance matrix which has the largest eigen-value. Eigen-vectors of course provide a direction, but not magnitude. Normalizing the eigen-vectors by the 1-norm, PCA then declares that:

$$z_{PCA} = \alpha z^1 + (1 - \alpha)z^2$$

The parameter $\alpha$ is determined such that $[\alpha (1 - \alpha)]'$ is parallel to the eigen-vector corresponding to the largest eigen-value.

By construction,

$$z_{PCA}^1 = \alpha p^1 z + (1 - \alpha)p^2 z + \alpha u^1 + (1 - \alpha)u^2$$

$$= (\alpha p^1 + (1 - \alpha)p^2) z + \alpha u^1 + (1 - \alpha)u^2$$

$$= p_{PCA}^1 z + \alpha u^1 + (1 - \alpha)u^2$$

The previously defined $p_{PCA}$ is the unidentifiable constant of proportionality which is trivially equal to $p$ if $p^1 = p^2 = p$. When trying to identify $p_{PCA}$, one may regress $y^1$ on $w^1$ and $z_{PCA}$. On the other hand, $z_{PCA}$ contains the residual in $y^1$, hence the
regression results will be plagued by the endogeneity problem. The rest of this paper focuses on how to best solve this endogeneity problem in two contexts: the simpler case where the dependent variable is not auto-regressive, and then the more involved case where the dependent variable is auto-regressive.

### 3.3.1 Non-Auto-Regressive Dependent Variable

To facilitate the discussion I will define the two following DGPs:

\[ y^1 = w^1 \beta^1 + p^1 z + u^1 \]  
\[ y^2 = w^2 \beta^2 + p^2 z + u^2 \]  

(3.12) \hspace{2cm} (3.13)

We will also assume \( w^1 \) and \( w^2 \) are exogenous random-variables which are orthogonal to \( z \). Hence we can consistently estimate \( \hat{\beta}^1 \) and \( \hat{\beta}^2 \) to calculate the following two proxies as residuals of the two regressions.

\[ \tilde{z}^1 = \hat{p}^1 z + \hat{u}^1 \]  
\[ \tilde{z}^2 = \hat{p}^2 z + \hat{u}^2 \]  
\[ \tilde{z}^{PC(1,2)} = PC(\tilde{z}^1, \tilde{z}^2) \]

One obviously cannot regress \( y^1 \) on \( w^1 \) and \( \tilde{z}^1 \). For similar reasons \( \tilde{z}^{PCA} \) cannot be used either. A seemingly viable alternative however may be regressing \( y^1 \) on \( w^1 \) and \( \tilde{z}^2 \). Although this solves the endogeneity problem, due to the error in \( \tilde{z}^2 \)'s approximation of \( z \), we will have the classical error-in-variable problem, hence \( \tilde{z}^2 \)'s coefficient will be biased.

The bias resulting from \( \tilde{z}^2 \)'s error can be corrected by using an instrument whose error is orthogonal to \( \tilde{z}^2 \)'s error. \( \tilde{z}^1 \)'s error is assumed orthogonal, but it can not be used as the instrument must also be exogenous. Inevitably, it seems like the factor's
coefficient cannot be identified unless there exists a third dependent variable.

\[ y^3 = w^3 \beta^3 + p^3 z + u^3 \]  

(3.14)

This will allow us to define:

\[ \tilde{z}^3 = \hat{p}^3 z + \hat{u}^3 \]

\[ \tilde{z}_{PC(1,2,3)} = PC(\tilde{z}^1, \tilde{z}^2, \tilde{z}^3) \]

Executing an instrumental-variable least-square (IVLS) regression, by regressing \( y^1 \) over \( w^1 \) and \( \tilde{z}^2 \), while using \( w^1 \) and \( \tilde{z}^3 \) as instruments results in an unbiased estimate of \( \frac{p^1}{p^2} \).33

For Figure 3.8, I construct equations 12, 13 and 14, with \( p^1 = p^2 = p^3 = 1 \).34

In the first step, I regress each \( y^i \) over \( z^i \) using OLS. I then document the standard error of \( \hat{\beta}^1 \) as well as the standard-deviation ratio statistic for \( \tilde{z}^1 \) and \( \tilde{z}_{PC(1,2,3)} \). In step two I regress \( y^1 \) over \( w^1 \) and \( \tilde{z}^2 \), using \( w^1 \) and \( \tilde{z}^3 \) as instruments and note the standard error of \( \hat{\beta}^1 \), the standard error of \( \frac{\hat{p}^1}{\hat{p}^2} \) as well as the two standard-deviation ratio statistics. Finally, I repeat step two and record the new results. The noted standard-error from step one, and two standard errors from step two, as well as the reiteration of step two, are graphed in Figure 3.8. The first column on the left shows results from a non-auto-regressive \( z \). In the second column from left \( z \) is defined as \( (1 - 0.5L)z_t = \epsilon_t \), where \( \epsilon_t \) is white-noise. The coefficient of auto-regression in the third column is 0.9 and in the forth column it is 0.99. Likewise in the first row \( w^i \) is non-auto-regressive. Its coefficient of auto-regression is 0.5, 0.9 and 0.99 in the second, third and forth rows respectively.

33Some papers in this literature call this the multiple-indicator method.

34The Monte-Carlo simulation was executed over 10,000 independent draws.
As evident from the first column, $\hat{\beta}^1$'s standard error does not improve from step one to step two. The reason is that as $\tilde{z}^2$, the regressor for $z$, consistently subtracts $z$ from the residual, it adds $u^2$ to $u^1$. In conjunction with that, the assumption in the DGP has been that $z$’s variance is the same as that of the error terms, as they are both standard normally distributed, so the variance of the regression residual is unchanged. In the second column, as $z$ is defined as $(1 - 0.5L)z_t = \epsilon_t$, I maintain the variance of $\epsilon_t$ to be the same as that of the errors, hence $z$ becomes more volatile. Consequently, from step one to step two, the variance of the residual is reduced by the more volatile $z$ and incremented by the less volatile $u^2$, hence the standard error of the estimator drops. This effect becomes much stronger in the third column, although there is still little advantage to repeating step two. This last reiteration however becomes worthwhile when $w$ and $z$ are both highly auto-regressive.

Figure 3.9 shows the standard deviation ratios. The solid line is \( \frac{\sigma(z - \tilde{z}^1)}{\sigma(z)} \), and the dotted line is \( \frac{\sigma(z - PC(1, 2, 3))}{\sigma(z)} \). As evident, the latter statistic consistently performs better than the former; reiteration however is of little help. Only when both $w$ and $z$ are highly persistent, does repeating step two improve the statistic.

Figure 3.10 and Figure 3.11 are similar to Figure 3.8 and Figure 3.9, with one difference. The difference is that there are five equations involved. The standard errors reported in Figure 3.10 are for $\hat{\beta}^1$ and $\hat{p}^{1}_{PC(3, 4, 5)}$, where the regression is an instrumental variable least square, regressing $y^1$ over $w^1$ and $PC(\tilde{z}^3, \tilde{z}^4, \tilde{z}^5)$, using $w^1$ and $\tilde{z}^2$ as instruments. In both of these figures, step two has been executed three times, as apposed to twice earlier.

The result repeats the same conclusions as Figure 3.8 and Figure 3.9, with one exception. As evident in the left most column of Figure 3.10, the standard error
Figure 3.8: Standard Errors For System of Three Equations

Coefficient of auto-regression of latent-factor is 0, 0.5, 0.9 and 0.99 in each column, from left to right, respectively. Coefficient of auto-regression of exogenous-regressor is 0, 0.5, 0.9 and 0.99 in each row, from top to bottom, respectively. Dashed line represents the standard-error of estimator of coefficient of exogenous-regressor, and dot-dashed line represents standard-error of coefficient of latent-factor.
Figure 3.9: Standard Deviation Ratio for System of Three Equations

Coefficient of auto-regression of latent-factor is 0, 0.5, 0.9 and 0.99 in each column, from left to right, respectively. Coefficient of auto-regression of exogenous-regressor is 0, 0.5, 0.9 and 0.99 in each row, from top to bottom, respectively. Solid line represents the standard-deviation ratio statistic for the $\tilde{z}_1$ proxy, and the dashed line represents the statistic for the $\tilde{z}_{PC(1,2,3)}$ proxy.
of $\hat{\beta}^1$ improves from step one to step two. The reason is that in step two, due to consistent estimation, the $z$ in $PC(\tilde{z}^3, \tilde{z}^4, \tilde{z}^5)$ subtracts the $z$ from the residual, while it adds $\sum_{i=3}^{5} \alpha_{i-2} u^i$, where $(\alpha_1, \alpha_2, \alpha_3)$ are the three components of the eigenvector with highest eigen-value, and $\sum \alpha_i = 1$ due to normalization. Hence $\text{var}(z)$ is larger than $\text{var}(\sum_{i=3}^{5} \alpha_{i-2} u^i) = \sum_{i=3}^{5} \alpha_i^2 \text{var}(u^i)$, if $\alpha_i < 1$, since by construction $\text{var}(z) = \text{var}(u^i)$. This result is significant because it explains why Monte-Carlo simulations suggest using $PC(z^3, z^4, z^5)$ as regressor and $z^2$ as instrument is more efficient than the other way around, where $z^2$ is a regressor and $PC(z^3, z^4, z^5)$ is an instrument. This result also further underlines PCA’s role in improving estimator efficiency in regressions.\textsuperscript{35}

The first important observation is that while in presence of three dependent variables $PC(\tilde{z}^1, \tilde{z}^2, \tilde{z}^3)$ provides the best approximation of $z$, it does not provide the best regressor for $z$. Indeed $\frac{p^1}{p^C(1,2,3)}$ cannot be identified, rather either of $\frac{p^1}{p^2}$ or $\frac{p^1}{p^3}$, depending in the choice of regressor, can be identified.

A second and important observation is that the residual from a regression with PCA generated regressors potentially has a smaller variance compared to variance of $\tilde{z}^i$, the residual from regressing the observable $y^1$ on the observable $w^1$. Consequently, the second-step IVLS regression could be more efficient with regards to $w^1$'s coefficient. Since the IVLS regression improves the efficiency of $w^1$'s estimator, the

\textsuperscript{35}PCA also seems to manipulate sample correlation between $u^i$'s in favor of improving estimator efficiency. This feature emerges in Monte-Carlo simulations executed on time-series of $T = 100$ and $T = 10,000$ observables. As number of observables increases, the marginal improvement from using a PCA generated regressor, as apposed to a single proxy $\tilde{z}^i$, diminishes. I hypothesize that as $T$ increases sample moments converge the population moments and sample correlations between $u^i$'s converge to zero, reducing PCA’s impact to be channeled solely through the mechanism discussed in the text above.
Figure 3.10: Standard Errors For System of Five Equations

The regressor used for the latent factor is PCA generated, taking three proxies as input. Coefficient of auto-regression of latent-factor is 0, 0.5, 0.9 and 0.99 in each column, from left to right, respectively. Coefficient of auto-regression of exogenous-regressor is 0, 0.5, 0.9 and 0.99 in each row, from top to bottom, respectively. Dashed line represents the standard-error of estimator of coefficient of exogenous regressor, and dot-dashed line represents standard-error of coefficient of latent-factor.
Figure 3.11: Standard Deviation Ratio Statistic for System of Five Equations

The regressor used for the latent factor is PCA generated. Coefficient of auto-regression of latent-factor is 0, 0.5, 0.9 and 0.99 in each column, from left to right, respectively. Coefficient of auto-regression of exogenous-regressor is 0, 0.5, 0.9 and 0.99 in each row, from top to bottom, respectively. Solid line represents the standard-deviation ratio statistic for the $\tilde{z}_1$ proxy, and the dashed line represents the statistic for the $\tilde{z}_{PC(1,2,3,4,5)}$ proxy.
approximation of \( z \), as represented by \( \tilde{z}^i \), will be improved so it is recommended to reiterate the procedure, specially where \( w \) and \( z \) are highly auto-correlated.

Third observation is that there has been made no assumption that would prevent \( w^i \) or \( z \) from being auto-regressive. In fact not only does the process remain consistent if either process shows persistence, the process also improves estimator efficiency.

Forth observation would be that at least three dependent variables are necessary to consistently estimate the coefficients using this method. What happens when 4 such dependent variables \( y^i \) are available? How can information resulting from \( \tilde{z}^4 \), as the 4th proxy of \( z \), be best used? While the classical approach has been to over-identify the instrumental-variable approach, through perhaps 2SLS or GMM, Monte-Carlo simulations suggest that PCA more effectively improves the standard-errors of the estimators. Furthermore, it turns out either of \( \tilde{z}^2 \), \( \tilde{z}^3 \), and \( \tilde{z}^4 \) can be used as instruments, with the other two combined to make a regressor. The choice of instruments will affect the standard-error proportional to its effect on the coefficient stemming from differences in \( p^{PC(2,3)} \), with \( p^{PC(2,4)} \) and \( p^{PC(3,4)} \). Hence for inference purposes it does not matter which is chosen as the t-value will remain the same.

In Figure 3.12 I repeat the exact same simulation behind Figure 3.10 but change the regressor. Instead of using \( PC(z^3, z^4, z^5) \) as the regressor and \( \tilde{z}^2 \) as the instrument, I use \( \tilde{z}^2 \) as the instrument and run a 2SLS where \( \tilde{z}^3 \), \( \tilde{z}^4 \), and \( \tilde{z}^5 \) are all used to over-identify \( \tilde{z}^2 \). Figure 3.12 is the table of graphs illustrating the standard-error of the estimated coefficients. The lighter green lines are repeats from Figure 3.10. In other words they are standard-errors resulting from a PCA based regression. The darker blue lines are results from a 2SLS regressions. It is quite evident that across both regressors, in each iteration, under all 16 combinations of regressor persistence,
the PCA method dominates the 2SLS. This does not mean PCA should always be regarded as more efficient than 2SLS. Rather when *a priori* information about the time-series implies a condition like above, 2SLS can be improved upon. In fact in the next section, I will show a case where 2SLS is more efficient than PCA.

Another common econometric method used in improving efficiency is FGLS. In Figure 3.13, I compare IVLS based results from a three-equation system, against FGLS. Figure 3.13 repeats results from Figure 3.8, and adds stars to show the standard error from regressing $y^1$ on $w^1$. Results shown in the first column from left show when the latent factor is not auto-correlated, hence the residual from the first-step is non-auto-correlated, and expectedly FGLS performs similar to IVLS. This does not imply FGLS is as good as IVLS when multiple proxies are available, as FGLS does not tell us $p$ or give an approximation of the latent factor. As we move to columns on the right, FGLS does outperform the first stage of IVLS; however, it cannot outperform the second-stage or its reiterations. Note the stars drawn for values 2 and 3 of the horizontal access are simply repeats of the first one, to facilitate comparison with the reiterations of the IVLS regression.

### 3.3.2 Auto-Regressive Dependent Variable

To understand why the previous section’s methodology is not immediately applicable to auto-regressive dependent variables let’s define the following DGP’s for $i = 1..4$: 36

$$y^i_t = \rho^i y^i_{t-1} + w^i_1 \beta^i + p^i z_t + u^i_t \tag{3.15}$$

36Since up to this point the temporal correlations were of no impact, the subscript $t$ was omitted. The introduction of $t$ in this section is simply to facilitate the conversation and does not imply any major change except for temporal dependency.
Figure 3.12: Comparison Between PCA Based Least Square and 2SLS, With Regards to Estimator Efficiency

The estimation uses one exogenous proxy as regressor and the other four to over-identify, using 2SLS. Result shown in dark-blue. Light-green represents results from PCA generated regressor. Coefficient of auto-regression of latent-factor is 0, 0.5, 0.9 and 0.99 in each column, from left to right, respectively. Coefficient of auto-regression of exogenous regressor is 0, 0.5, 0.9 and 0.99 in each row, from top to bottom, respectively. Dashed line represents the standard-error of estimator of coefficient of exogenous regressor, and dot-dashed line represents standard-error of coefficient of latent-factor.
Figure 3.13: Comparison Between IVLS and FGLS in Three Equation System. Star: $\beta$, Dash: $\beta$, Dot-Dash: $p$.

Dark-blue lines represent results from an IVLS regression with three dependent-variables. Star shows the standard-error of the estimator of the coefficient of exogenous regressor, when a twice iterated FGLS is used. The three stars in each chart are repeats from the second-iteration, facilitating comparison with different iterations of IVLS. Dashed line represents the standard-error of estimator of coefficient of exogenous regressor, and dot-dashed line represents standard-error of coefficient of latent-factor.
If one were to regress \( y_i^t \) on \( y_{i-1}^t \) and \( w_i^t \), the residual would be a proxy for \( z_t \) only if \( z_t \) was not auto-correlated. On the other hand, if \( z_t \) was an auto-correlated process, \( y_{i-1}^t \) would be correlated with \( z_{t-1} \) and hence the first-step regression would result in a biased estimation of \( \rho \) which consequently biases our proxy for \( z_t \).

While the previous method now only holds if \( z_t \) is not auto-correlated we need a method for the more general case. Here as well, I will resort to the instrumental-variable method. If we regress \( y_i^t \) over \( y_{i-1}^t \) and \( w_i^t \) while using \( w_{i-1}^t \) and \( w_i^t \) as corresponding instruments, we will get consistent though not particularly efficient estimators. This serves as a first step to estimate the proxies as:

\[
\tilde{z}_i^t = y_i^t - \hat{\rho} y_{i-1}^t - w_i^t \hat{\beta} = \hat{p} z_t + \hat{u}_i^t \tag{3.16}
\]

In the next step, we can regress \( y_1^t \) over \( y_{1-1}^t \) and \( w_1^t \), as well as \( PC(z^3, z^4) \), using \( y_{1-1}^t, w_1^t \) and \( \tilde{z}_2^t \) as instruments. Note now that we do have a proxy for \( z_t \), we can use \( y_{t-1}^t \) as its own instrument. Simulation results suggest re-estimating \( \tilde{z}^i \) using newly estimated coefficients and re-executing the second step of IVLS improves efficiency. Indeed 2 to 3 iterations seem to yield optimal result in most cases.

In order to better understand IVLS’s behavior in the auto-regressive context we can look at Figure 3.14 though Figure 3.19. For the production of these graphs, I have generated 10,000 independent versions of a system of five equations, each like equation 32. Four different values of \( \rho^i \) was used, at 0, 0.5, 0.9 and 0.99 respectively, represented in each row of the figures. The \( \beta^i \) coefficient, as well as \( p^i \) were fixed at 1. The exogenous regressor \( w_i^t \) was AR(1) with coefficient of 0.9 in all cases. The latent factor \( z_t \) was generated with four different autoregressive coefficients, 0, 0.5, 0.9 and 0.99 respectively, represented in each column of the figures. The five
\( u^i \)'s were independently drawn from standard normal distributions, as idiosyncratic disturbances.

Figure 3.14 starts by assessing consistency and shows the three estimated coefficients in each iteration. In the first column from left, where the latent factor is non-auto-regressive, \( \beta \) and \( p \) are consistently estimated at one, and the dots show an unbiased estimation of \( y \)'s persistence. The forth column from left is where the latent factor is almost a random-walk. The top row shows all three estimators are biased until the last iteration. The third row exhibits the need for even more iterations when \( y \) is so persistent. The need for further iterations is ever more dire in the case where \( y \) is also an almost-random-walk, hence the second step is iterated twenty times, instead of three. The dots, showing \( y \)'s persistence, converge to the unbiased level rather quickly. The latent factor's coefficient, represented by the dot-dash line converges much slower. The exogenous regressor’s coefficient however, represented by the dashed line, does not converge to 1, and is indeed biased.

Figure 3.15 compares PCA’s performance with 2SLS. In Figure 3.14, the regressor used to proxy the latent factor was generated as the PCA of 3 proxies. In Figure 3.15 however, we will use one proxy and use the rest of to over-identify the regression. The light-green lines are repeating results from Figure 3.14. The dark-blue lines are average estimated coefficients using 2SLS. The three columns to the left show that PCA-based regressions have smaller bias initially, and reach unbiased levels more quickly; but after three iterations of step-two, there is not much difference remaining between the two methods. In the forth column however, where the latent factor is highly persistent, at 0.99, 2SLS performs better. In the top two rows of that column, the two methods end up equally consistent after multiple iterations, but not in the
bottom two rows. In the third row, where \( y \)’s persistence coefficient is 0.9, 2SLS is unbiased after 4 iterations, but PCA based regression requires more iterations. When both \( z \) and \( y \) are almost-random-walks (column 4, row 4), the persistence level of \( y \) is initially overestimated, as 1, before it finally converges to the unbiased level. The latent-factor’s coefficient is also unbiased, similar to PCA-based, but reaches the unbiased level much faster. 2SLS and PCA-based regressions both make biased estimations of the exogenous regressor’s coefficient.

Figure 3.16 shows the standard-errors of the estimated coefficients. The first two columns on the left show the benefit of executing the second step in improving the efficiency of \( \hat{\beta} \) and \( \hat{\rho} \). The third column illustrates the benefit or reiterating the second-step. Figure 3.17 repeats results from Figure 3.16 in light-green color, and adds the standard-errors from the 2SLS regression in dark-blue. While the first three columns make the advantage of PCA-based regressions obvious, the forth column again emphasizes the benefit of over-identification when the latent-factor is an almost-random-walk. In the forth column, PCA still outperforms 2SLS in the first row, while the result is mixed in the second row. The forth row, showing results from a 20 iteration regression, shows dark-blue lines which are consistently below the light-green ones, pointing at 2SLS superiority.

Figure 3.18 and Figure 3.19 graph the standard-error ratio statistic for the approximated latent factor. The dashed line is the statistic resulting from comparing the true latent factor against \( \tilde{z}^1 \), and the dot-dashed line shows the statistic when the true latent factor was compared against the PCA of all 5 proxies. Looking at Figure 3.19, the first three columns show PCA to be negligibly outperforming 2SLS, while the forth column, consistent with prior observations, declares 2SLS as the more
efficient. Of course its noteworthy that after many iterations, the top three rows find 2SLS and PCA on par. 2SLS’s efficiency becomes significantly pronounced only when both the latent-factor, and the dependent-variable are AR with coefficient of 0.99.

In contrast to the previous section, note that the exogenous regressor, \( w^i \), plays a critical role in this section. Whereas before, in absence of \( w^i \), we could have used \( y^i = p^i z + u^i \) perfectly well to approximate \( z \) and run a regression; now in the case where \( y^i_t = \rho y^i_{t-1} + p^i z_t + u^i_t \) we cannot approximate a proxy for \( z_t \) anymore.

### 3.4 VAR

Perhaps the most celebrated empirical method is the VAR. The procedure described above is generalizable to VAR’s. Suppose we have a system of equations looking like:

\[
y^i_t = \sum_{j=1}^{4} (\rho^j y^j_{t-1} + w^j_t \beta^j) + p^i z_t + u^i_t
\]

(3.17)

We can repeat a procedure similar to that explained above. However, it is noteworthy that for every \( y^j_{t-1} \) in the equation, we need an exogenous regressor, \( w^j_t \), to use as an instrument in the first step. In other words, just one \( w^i_t \) per equation may not suffice anymore for identification.

### 3.4.1 Factor Augmented VAR

The econometric literature’s interest in PCA has primarily laid in using the PCA enhanced proxy as a time-series augmented to a VAR to improve estimation results. Where above, the latent factor was just an explaining-variable for the \( y^i \)’s, in this FAVAR approach, the latent factor is meant to be augmented to the vector of \( y^i \)’s and receive feedback from their lagged realizations.
Figure 3.14: Coefficients from PCA-Based Regression. Dot: $\rho$, Dash-Dot: $\rho$, Dashed: $\beta$

PCA-generated regressor proxies for the latent-factor, where the dependent variable is AR(1). The dots represent the estimated coefficient of auto-regression. Dashed lines represent coefficient of exogenous regressor. Dot-dashed lines represent coefficient of latent-factor.
Figure 3.15: Coefficients from 2SLS Regression. Dot/ρ, Dash-Dot/p, Dashed/β

Light-green lines are repeats from Figure 3.14. Dark-blue represents results from 2SLS regression. The dependent variable is AR(1). The dots represent the estimated coefficient of auto-regression. Dashed lines represent coefficient of exogenous regressor. Dot-dashed lines represent coefficient of latent-factor.
Figure 3.16: Standard-Errors from PCA-Based Regression. Dot: $\rho$; Dash-Dot: $p$; Dashed: $\beta$. PCA-generated regresor proxies for the latent-factor, where the dependent variable is AR(1). The dots represent the standard-error of estimated coefficient of auto-regression. Dashed lines represent standard-error of coefficient of exogenous regressor. Dot-dashed lines represent standard-error of coefficient of latent-factor.
Figure 3.17: Standard-Errors from 2SLS Regression. Dot: $\rho$, Dash-Dot: $\beta$, Dashed: Light-green lines are repeats from Figure 3.16. Dark-blue represents results from 2SLS regression. The dependent variable is AR(1). The dots represent the standard-error of estimated coefficient of auto-regression. Dashed lines represent standard-error of coefficient of exogenous regressor. Dot-dashed lines represent standard-error of coefficient of latent-factor.
Figure 3.18: Standard-Deviation-Ratio Statistic from PCA-Based Regression.

PCA-generated regressor proxies for the latent-factor, where the dependent variable is AR(1). Dashed line represents the standard-deviation ratio statistic for the $\tilde{z}_1$ proxy, and the dot-dashed line represents the statistic for the $\tilde{z}_{PC(1,2,3,4,5)}$ proxy.
Figure 3.19: Standard-Deviation-Ratio Statistic from 2SLS Regression

Light-green lines are repeats from Figure 3.18. Dark-blue represents results from 2SLS regression. The dependent variable is AR(1). Dashed line represents the standard-deviation ratio statistic for the $\tilde{z}^{1}$ proxy, and the dot-dashed line represents the statistic for the $\tilde{z}^{PC(1,2,3,4,5)}$ proxy.
Let $Y = [y^1 \cdots y^I]'$ be a vector of $I$ observable time-series, and let $W = [w^1 \cdots w^J]'$ be a vector of $J$ observable regressors. Further assume $F_t$ is a latent time-series reflecting the state of the economy, regarded by the business cycle literature as the diffusion index. For example, one can think of $F_t$ as total-factor-productivity, level of economic activity, or any factor of penchant. Assume the economy’s data is drawn from the following process.

$$Y_t = Y_{t-1} \Phi + W_t \Lambda + F_t \rho + U_t \tag{3.18}$$

$$F_t = Y_{t-1} \phi + W_t \lambda + z_t \tag{3.19}$$

$\Phi$ is an $I \times I$ matrix, $\Lambda$ is an $I \times J$ matrix, and $\rho$ is an $I \times 1$ matrix and $\lambda$ is a $J \times 1$ matrix. $U_t$ is a $I \times 1$ vector and $e_t$ is a scalar time-series. All variables satisfy all assumptions necessary for OLS to work, had $F_t$ been observable.

The FAVAR approach involves stacking the $I$ components in $Y_t$ over $F_t$, to run a more efficient VAR. The system of equations proclaims all observable macroeconomic factors partially explain, and are explained by, the state of the economy as represented by $F_t$. But if $F_t$ is a function of lagged observables, what makes it unobservable? The answer is $z_t$ together with the parameters. So the real latent factor is not $F_t$, rather it is $z_t$. Indeed we can write the two equations above as:

$$Y_t = Y_{t-1}(\Phi + \phi \cdot \rho') + W_t(\Lambda + \lambda \cdot \rho') + p \cdot z_t + U_t \tag{3.20}$$

The common factor in all $I$ components of $Y_t$ is $z_t$. If we could get an estimation of the unobservable factor $z_t$, then we could do better than just running a VAR of $Y_t$ over its lags and $W_t$. If the implied equation above has sufficiently many exogenous regressors in $W_t$, we can apply the aforementioned procedure by developing proxies.
and instruments for the latent factor, and get more efficient estimates of the $I \times I$ matrix $\Phi + \phi \cdot p'$ and the $J \times I$ matrix $\Lambda + \lambda \cdot p'$, than we previously could.

Another application for FAVAR is counter-factual studies. In such practices, an econometrician would like to do a what-if study, manipulating an exogenous regressor and study the impact on a time-series of interest. While this could be done in VAR, including latent-factors could certainly improve the results. Within the Stock and Watson structure, one has to estimate the factor, $F_t$, find the effect of the exogenous regressor on the factor, and the effect of both on the explained-variable. A counter-factual exogenous regressor will be fabricated, then the counter-factual latent-factor will be approximated and finally the counter-factual explained variable will be estimated. On the other hand, in the proposed procedure the latent-factor is defined as $z_t$, which is uncorrelated with the exogenous regressor. This cuts one step in the middle, as there is no need to estimate a counter-factual latent-factor. After determining the linear relationships, the econometrician can go straight from the counter-factual exogenous regressor to the counter-factual explained-variable.

### 3.5 Summary

Principal Components Analysis is quite a powerful tool for enhancing proxies. But the proxy with reduced noise should not be immediately considered as a better regressor. Theoretically, as $N \to \infty$, it can be considered as a better regressor, but empirical tests, cited earlier, have shown otherwise. When $N$ is not large enough, which it often is not, endogeneity could make PCA quite counter-productive, unless instruments are used. Moreover by construction PCA creates an error-in-variable
problem which leads to biased estimation. Instruments can also serve a purpose in this case.

This paper describes the problem and its solution, and utilizes Monte-Carlo based results to prescribe best methods for incorporating all available information. The major contributions of this paper are: 1) presenting a procedure which given sufficient time-series can be superior to FGLS; 2) presenting the circumstances when the procedure is superior to 2SLS; 3) presenting an instrument-based solution for cases where latent factors are hidden in auto-regressive time-series; 4) last but not least, providing a comparison between the proposed method and Stock and Watson’s method, in three different contexts.
Chapter 4: CONTRIBUTIONS AND FUTURE WORK

A considerable portion of the asset pricing literature considers the demand schedule for asset prices to be perfectly elastic (flat). As argued, asset prices are determined using information about future payoff distribution, as well as the discount rate; consequently, an asset would be priced independent of its available supply. Furthermore, such a flat demand curve is considered to be a consequence of the Efficient Market Hypothesis. My dissertation evaluates and questions the factuality of these assertions. I approach this problem from both an empirical and a theoretical perspective. The general argument is that asset prices do respond to supply-shocks; and changes in aggregate demand, stemming from preference changes, new international investments, or quantitative easing by the Fed, can result in price changes. Hence, asset prices are determined by both demand and supply factors.

In the first essay, Downward Sloping Asset Demand: Evidence from the Treasury Bills Market, I reported on my empirical study which establishes the existence of a downward sloping demand curve (DSDC) in the T-bill market. I showed that the events of T-bill auction and T-bill maturity lead to persistent abnormal yields, indicating the significance of supply in pricing T-bills. The asset pricing literature contains multiple empirical evidence of a DSDC, but such evidence has generally been considered inconclusive. Alternative hypotheses have been developed, such as
the price-pressure hypothesis and the information hypothesis, to explain the empirical
evidence in line with a flat demand curve. My essay, however, presents evidence from
a liquid, risk-free market with frequent uninformative supply shocks. Due to these
unique features of the T-bill market, none of the alternative hypotheses apply, thus
the essay presents a particularly conclusive evidence for the existence of DSDC.

The empirical method is based on an event study at the auction and maturity
events. The regression errors are modeled as an autoregressive process with gener-
alized GARCH innovations, to account for the pronounced heteroskedasticity, serial
correlation and leptokurticity in the data. I investigated not only the supply effect,
but also its persistence. The standard method for persistence estimation is examining
the cumulative-abnormal-response; however, due to the high frequency of events, this
method is not applicable to the T-bills market. In the first essay, I developed a simple
and intuitive method to infer the persistence of the effect. I found that about half
of the auction effect persists, whereas the maturity effect is completely transitory. In
the second essay, Asset Pricing: Inelastic Supply, I examined the theoretical issues
concerning a downward sloping demand curve. I began by clarifying a common con-
fusion in the literature, namely, that many asset pricing models imply a flat demand
curve. I showed that the prominent asset pricing models, including CAPM (Capital
Asset Pricing Model), APT ( Arbitrage Pricing Theory) and CCAPM (Consumption
Capital Asset Pricing Model), all have an underlying DSDC. I further showed that,
while these models imply the relevance of supply, they are inconvenient as a vehicle
for the estimation and analysis of the DSDC in the data. For those purposes, I devel-
oped an asset pricing framework based on the stochastic discount factor framework,
specifically designed with a DSDC at its heart. I ended the essay with a discussion of the frameworks implications and applications.

The two essays together present evidence for DSDC, argue for potential efficiency of an asset market with DSDC, and present a theoretical framework for future research in modeling DSDC in the asset market.

The third essay was on Econometrics. I presented the power of Principal Components Analysis which made it enticing as a regressor generator. I then discussed the current literature on the topic and exposed the resulting bias. Comparing the prominent methods with mine I made the case for the incorporation of instrumental-variable least-square into FAVAR estimations.
Appendix A: ARBITRAGE EXAMPLE

In an economy with a linear asset pricing function, let the market contain an asset with payoff distribution $A$ which is driven by two factors $T$ and $U$ in the following form: $A \equiv T + U$. Obviously, the price of $A$ is the sum of $T$ and $U$’s price. Let asset $B$ be defined as $B \equiv U + V$, where $T$, $U$ and $V$ are three independently distributed factors. Let there be 100 shares of each asset, both facing a DSDC, and the market be at equilibrium. Doubling the number of shares of asset $A$ in the market will reduce its price, but doubling $A$ means increasing the number of $T$’s from 100 to 200, and the number of $U$’s from 200 to 300 while the number of $V$’s remains 100. This means one of the factors in $B$ is facing an increased supply as well and hence will be priced lower. This reduces the price of $B$, but this price change will be different from that of $A$, and that is expected as the two assets are imperfect substitutes. In this way, a supply shock to an asset will affect all correlating assets.

Next example shows how a supply shock to an asset can affect a portfolio which does not even contain the shocked asset, but simply correlates with it. To have a market with perfect substitutes, add asset $C \equiv T$ and asset $D \equiv V$ to the market, with 100 shares of each. Create a portfolio which is one unit long on $B$, one unit long on $C$ and one unit short on $D$. This portfolio, call it $E$, is a perfect substitute for $A$. Doubling the number of $A$’s from 100 to 200 means asset $C$ experiences a supply
shock, while asset $D$ remains unchanged. The underlying factors in portfolio $E$ have faced the exact same supply shock as the factors in asset $A$; therefore, $E$ will face a price change just as asset $A$ would.
Appendix B: TIME EVOLUTION OF SUPPLY-EVENT RESPONSE

When running a regression over 6500 data points spread over a 26 year period, one wonders if the coefficients are significant in sub-periods too. In order to study sub-periods, a similar regression as Table 1.6 will be executed, but only over the first four years of the data.\(^{37}\) Similar to Table 1.6, the estimation window is the four days prior to the event window.

After the first regression is run over approximately four years (1000 daily observations), the observation window will be shifted by a day, and the same regression will be repeated. This procedure is repeated until 5500 regressions are executed. Finally, the resulting coefficients are plotted with two standard errors above and below. Figures B.1, B.2 and B.3 show the announcement, auction and maturity coefficients, respectively. Each figure contains the graph of a coefficient under the dummy regression, as well as the graph under the structured regression. Each point is the coefficient over four years, where the first point on the left is the coefficient for a regression over Jan 4th 1982 through Jan 7th 1986. The second point is the coefficient for a regression over Jan 4th 1982 through Jan 7th 1986. The second point is the coefficient for a regression over Jan 4th 1982 through Jan 7th 1986.

\(^{37}\)Due to computational intensity, the graphs in this section are estimated using GLS instead of GARCH. Refer to footnote 15. The one major difference is that the Tuesday dummy will be removed because in certain sub-periods it is colinear with the Announcement dummy. To remain consistent throughout this sub-section, all regressions will be run over Monday, Wednesday, Friday, Announcement, Auction and Maturity regressors.
Figure B.1: Time evolution of the announcement coefficient
Figure B.2: Time evolution of the auction coefficient
Figure B.3: Time evolution of the maturity coefficient
over Jan 5th 1982 through Jan 8th 1986, and so on, until the last point which is the coefficient for the duration starting Jan 5th 2004 until 1000 business days later, Dec 31st 2007.

Figure B.1 shows the evolution of the announcement-day abnormal-yield as inferred by both the dummy based and structured regression. The dummy based regression estimates larger standard-errors for the earlier periods whereas the structured regression estimates larger standard-errors for the more recent periods. Nonetheless, both estimations agree on periods with statistically significant announcement-day effects. It is noteworthy that from around mid-1998 (about one year after the sudden regime change in unexpected volume) the announcement-day event has completely vanished. This explains our observation in the announcement column of Table 1.5, implying the unexpected component of an auction had nothing to do with a change in the announcement-event’s effect. In fact, from mid-1997 through mid-1998, we observe increased uncertainty in an announcement-event’s response. What happened around mid 1998 remains an unanswered question for this paper.

Figure B.2 however shows that past mid-1997 the auction event becomes considerably more significant than previous periods. This matches our expectation in presence of increased uncertainty. Till the end, this change stays strong, but of course the empirical experiment in Table 1.5 reveals that most of this increase is transitory. Luckily, the time-evolution study can give us the benefit of understanding the time-evolution of the temporary component as well. If an increase in the size of unexpected component of auctions increases abnormal-yield, we expect larger substitution effects earlier on, but since the size of the unexpected component stabilizes during the post April-1997 period, we expect the temporary component to gradually subside. This actually
Figure B.4: Time evolution of the temporary component of auction coefficient

does match information contained in the data, as Figure B.4 reveals. This figure contains the same graph from Figure B.2 with the auction coefficient from a regression with shifted estimation window overlaid. Since the four-standard-error band of this coefficient (showing temporary component), for visual purposes, was the same as that of the regression with estimation window preceding the event, the band is omitted from the graph. The graph clearly shows that following the April 1997 regime change, the temporary component rose considerably, just to later subside, primarily leaving the permanent component of the event response. In fact after 2002, the temporary component becomes statistically insignificant.
Figure B.3 indicates that maturity had been significant since early 90’s but during the 80’s the event of maturity had considerably large standard-errors. This event has had its strongest effect during the turn of the century.

All in all, the three events have been statistically significant during most of the 26 year period, although their effect has changed over time. Announcement event is one which has lost significance starting around mid 1998, but I do not know of the potential cause. On the other hand, abnormal-yield due to auction events has indeed systematically changed following the April 1997 increase of the unexpected component of auction volume.
Bibliography


Belviso, Francesco and Fabio Milani. “Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy”. *Topics in Macroeconomics*, 6, Dec 2006.


