Essays in Financial Econometrics

by

Emre Kocatulum

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

© Emre Kocatulum, MMVIII. All rights reserved.

The author hereby grants to MIT permission to reproduce and
distribute publicly paper and electronic copies of this thesis document
in whole or in part.

Author ............... 

Department of Economics

July 25, 2008

Certified by......

Victor Chernozhukov

Associate Professor

Thesis Supervisor

Certified by......

Whitney Newey

Jane Berkowitz and Dennis Carton Professor

Thesis Supervisor

Accepted by ..................

Peter Temin

Elisha Gray II Professor of Economics

Chairman, Departmental Committee on Graduate Studies
Abstract

Chapter 1 is the product of joint work with Ferhat Akbas and it provides a behavioral explanation for monthly negative serial correlation in stock returns. For the first time in the literature, this work reports that only low momentum stocks experience monthly negative serial correlation. Using a recently collected dataset, this finding provides the basis for a behavioral explanation for monthly negative serial correlation.

Chapter 2 uses mean squared error (MSE) criterion to choose the number of instruments for generalized empirical likelihood (GEL) framework. This is a relevant problem especially in financial economics and macroeconomics where the number of instruments can be very large. For the first time in the literature, heteroskedasticity is explicitly modelled in deriving the terms in higher order MSE. Using the selection criteria makes GEL estimator more efficient under heteroskedasticity.

Chapter 3 is the product of joint work with Victor Chernozhukov and Konrad Menzel. This chapter proposes new ways of inference on mean-variance sets in finance such as Hansen-Jagannathan bounds and Markowitz frontier. In particular standard set estimation methods with Hausdorff distance give very large confidence regions which are not very meaningful for testing purposes. On the other hand confidence regions based on LR-type statistic and wald type statistic provide much tighter confidence bounds. The methodology is also extended to frontiers that use conditional information efficiently.
Acknowledgments

I thank Victor Chernozhukov for his guidance, support and advice. During our joint project which became a part of this thesis I have learned a lot from him both in econometrics and generally, in doing research.

I thank Whitney Newey for leading me into fruitful areas of econometrics. Without his guidance and vision it would have been much more difficult to identify problems in econometrics that represent good research opportunities.

I thank Ferhat Akbas for supporting me during frustrating moments of the graduate program. Without ever coming together, our exchanges of ideas over the phone resulted in a part of this thesis. I thank Konrad Menzel for intellectually stimulating discussions.

I thank my roommate Joao Leao for his support especially during the initial years of my graduate studies. I thank my other roommate Frantisek Ricka for his moral support during times when things did not look very uplifting. I thank Mustafa Sabri Kilic for his moral support during my five years at MIT.

I thank MIT Department of Economics for financially supporting me during my studies.

I thank to my family for constantly supporting me. I dedicate this thesis to my father and mother.
## Contents

1 Momentum, Monthly Reversion and Theories of Investor Sentiment 13
   1.1 Introduction .................................................. 14
   1.2 Data and Preliminary Results ................................ 18
      1.2.1 Data Description .......................................... 18
      1.2.2 Preliminary Results ....................................... 19
   1.3 Monthly Reversion of High and Low Momentum Stocks ........ 19
      1.3.1 Fama-MacBeth Regressions .............................. 20
      1.3.2 Portfolio Returns ......................................... 21
   1.4 News and No News ............................................. 22
   1.5 Response to News: Overreaction of Low Momentum and ... 25
      Under reaction of High Momentum Stocks .................... 25
   1.6 Controlling for Various Cross-Sectional Effects and Robustness Checks 27
      1.6.1 Illiquidity ................................................. 28
      1.6.2 Turnover .................................................. 29
      1.6.3 Risk Factor Loadings of Monthly Contrarian Strategy Profits 31
      1.6.4 January Effect and Sub period Analysis ............... 32
   1.7 Conclusion .................................................... 32
   1.8 Appendix ........................................................ 34

2 Number of Instruments in GEL Estimation with Heteroskedasticity 47
   2.1 Introduction .................................................... 47
   2.2 The GEL Estimator ............................................. 49
3 Inference on Sets in Finance

3.1 Introduction ................................................. 85

3.2 Estimation and Inference Results ............................... 89
  3.2.1 Basic Constructions ........................................ 89
  3.2.2 A Basic Limit Theorem for LR and W statistics ........... 91
  3.2.3 Basic Validity of the Confidence Regions ................ 95
  3.2.4 Estimation of Quantiles of LR and W Statistics by Bootstrap
             and Other Methods ....................................... 96

3.3 Empirical Example .......................................... 100

3.4 Conclusion .................................................. 101

3.5 Appendix .................................................... 102
  3.5.1 Figures .................................................. 102
  3.5.2 Proofs .................................................. 108
List of Figures

3-1 Estimated HJ Bounds ........................................ 102
3-2 Estimated HJ Bounds and Bootstrap Draws .................. 103
3-3 90% Confidence Region using LR Statistic .................... 104
3-4 90% Confidence Region using Unweighted LR Statistic ........ 105
3-5 90% Confidence Region using Unweighted W Statistic (H-Distance) . 106
3-6 90% Confidence Region using Weighted W Statistic ............ 107
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary Statistics</td>
<td>34</td>
</tr>
<tr>
<td>1.2</td>
<td>Monthly Reversion and Momentum: Portfolio Approach</td>
<td>35</td>
</tr>
<tr>
<td>1.3</td>
<td>Interaction of Momentum and Monthly Reversion:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fama-MacBeth Regressions</td>
<td>36</td>
</tr>
<tr>
<td>1.4</td>
<td>Interaction of Momentum and Monthly Reversion: Portfolio Approach</td>
<td>37</td>
</tr>
<tr>
<td>1.5</td>
<td>Interaction of Momentum and Monthly Reversion for News and No News Stocks</td>
<td>38</td>
</tr>
<tr>
<td>1.6</td>
<td>Momentum and Illiquidity</td>
<td>39</td>
</tr>
<tr>
<td>1.7</td>
<td>Momentum and Turnover</td>
<td>40</td>
</tr>
<tr>
<td>1.8</td>
<td>Time Series Regression of Contrarian Strategy Profits on Fama-French Factors</td>
<td>41</td>
</tr>
<tr>
<td>1.9</td>
<td>January Effect and Sub period Analysis</td>
<td>42</td>
</tr>
<tr>
<td>2.1</td>
<td>CUE Estimators: Small Sample, Weak Endogeneity and Homoskedasticity</td>
<td>76</td>
</tr>
<tr>
<td>2.2</td>
<td>CUE Estimators: Small Sample, Weak Endogeneity and Heteroskedasticity</td>
<td>76</td>
</tr>
<tr>
<td>2.3</td>
<td>CUE Estimators: Small Sample, Medium Endogeneity and Homoskedasticity</td>
<td>77</td>
</tr>
<tr>
<td>2.4</td>
<td>CUE Estimators: Small Sample, Medium Endogeneity and Heteroskedasticity</td>
<td>77</td>
</tr>
<tr>
<td>2.5</td>
<td>CUE Estimators: Small Sample, Strong Endogeneity and Homoskedasticity</td>
<td>77</td>
</tr>
</tbody>
</table>
2.6 CUE Estimators: Small Sample, Strong Endogeneity and
   Heteroskedasticity ........................................... 78
2.7 CUE Estimators: Large Sample, Weak Endogeneity and
   Homoskedasticity ............................................. 78
2.8 CUE Estimators: Large Sample, Weak Endogeneity and
   Heteroskedasticity ............................................. 78
2.9 CUE Estimators: Large Sample, Medium Endogeneity and
   Homoskedasticity ............................................. 79
2.10 CUE Estimators: Large Sample, Medium Endogeneity and
    Heteroskedasticity ........................................... 79
2.11 CUE Estimators: Large Sample, Strong Endogeneity and
   Homoskedasticity ............................................. 79
2.12 CUE Estimators: Large Sample, Strong Endogeneity and
    Heteroskedasticity .......................................... 80
2.13 Instrument Selection Based on MSE ........................ 81
Chapter 1

Momentum, Monthly Reversion and Theories of Investor Sentiment

We document that the negative first-order serial correlation (reversion) in monthly stock returns is not a market-wide phenomenon as claimed by Jegadeesh [1990]. Low momentum stocks show strong negative serial correlation whereas high momentum stocks do not show any significant serial correlation. A strategy that utilizes this predictability in low momentum stocks provides monthly returns of 1.76% during 1980-2006. Using data from Chan [2003], we analyze the role of reaction to news in this finding. Low momentum stocks over react to news announcements, whereas high momentum stocks under react. Therefore, the correction to the news driven returns in low momentum stocks partially explains the monthly reversion of low momentum stocks. Our finding is supportive of a version of the investor sentiment model developed by Daniel et al. [1988].

1This Chapter is the product of joint work with Ferhat Akbas.
1.1 Introduction

Based on Fama [1970], weak-form market efficiency hypothesis is satisfied only if stock returns cannot be predicted using historical price data. Bondt and Thaler [1985] is one of the first of a series of important papers that document stock return predictability which is based on past stock performance. For three to five year intervals they document a significant negative serial correlation for stock returns. Jegadeesh and Titman [1993] document positive serial correlation when stock returns are measured for three to twelve month intervals. This predictability is called momentum. For even shorter intervals of one week and one month, Lehmann [1990] and Jegadeesh [1990] document significant negative serial correlation.

Since these predictabilities appear to contradict the efficient market hypothesis, a large number of studies have tried to explain these findings. This literature can be divided under three headings. Explanations based on risk, explanations based on behavioral biases and studies that show that these predictabilities cannot be used to create profitable strategies due to implementation difficulties such as illiquidity and bid-ask spread.

The main idea of risk based explanations is that the abnormal profits that can be generated using any of the significant predictabilities are actually nothing more than a reward for taking certain nondiversifiable risks. Therefore these strategies do not constitute a free lunch and hence do not contradict with the efficient market hypothesis. For example Chan [1988], Ball and Kothari [1989] and Zarowin [1990] all provide risk based explanations for the market anomaly documented by Bondt and Thaler [1985]. Fama and French [1996] develop three risk factors to control for risks assumed by various strategies. They document that with the exception of momentum anomaly most of the reported market anomalies disappear once the risks assumed by the underlying strategies are taken into consideration.²

A number of behavioral models are developed to explain momentum finding of

²This view is contested by Chopra et al. [1992].
³Fama and French [1996] do not investigate short term negative serial correlation as found by Lehmann [1990] and Jegadeesh [1990]. In our study, we show that the three Fama-French factors cannot explain away the profits generated by monthly negative serial correlation.
Jegadeesh and Titman [1993] and long run negative serial correlation finding of Bondt and Thaler [1985]. Daniel et al. [1988] (hereafter, DHS) presents a model with investor over confidence about private signals. The result is an under reaction to public news and over reaction to private signals. Investors in Barberis et al. [1998] (BSV) believe in either contrarian or trend following forecasting. The resulting equilibrium returns tend to continue in the short term and revert in long term. On the empirical side, Chan et al. [1996] shows that momentum result of Jegadeesh and Titman [1993] can be explained mostly by under reaction to information such as earnings announcements. Chan [2003] uses a news dataset to distinguish returns driven by public news and returns unaccompanied by public news. His results imply that stocks under react to public news which is in line with DHS model. Zhang [2006] documents that stocks with greater information uncertainty show stronger momentum which supports the behavioral explanations of momentum anomaly.

Noticeably, the risk based and behavioral literature mostly focuses on explaining 6-12 month momentum and longer term reversion. In contrast, a significant portion of the literature about bid-ask spread and illiquidity focuses on weekly and monthly negative serial correlation of Lehmann [1990] and Jegadeesh [1990]. This is not surprising as the high frequency trading suggested by these short term predictabilities makes them more vulnerable to transaction cost and illiquidity related criticisms. Therefore by showing the lack of profitability for these weekly and monthly strategies, one can reconcile these short term predictabilities with Jensen [1978] definition of market efficiency and Rubinstein and Stephens [2001] definition of minimally rational markets.\(^4\)

The literature has been more successful in contesting the profitability of weekly negative serial correlation compared to the profitability of monthly negative serial correlation. According to Jegadeesh and Titman [1995b], trading profits documented by Lehmann [1990] are not significant once the inventory risk is taken into account. Jegadeesh and Titman [1995b] claims that their results can be extended to monthly

\(^4\)Both Jensen’s and Rubinstein’s definition imply that if it is not possible to make profits from a predictability, than this predictability cannot be considered a market inefficiency.
returns, however their paper does not include an analysis for monthly returns. Similarly, Ball et al. [1995] and Conrad et al. [1997] finds that weekly contrarian profits are not robust to bid-ask bounce. Finally, Avramov et al. [2006] suggest that weekly and monthly negative serial correlations are illiquidity driven. However, their case is more convincing for weekly returns than for monthly returns.

Given the inadequate explanations of risk and bid-ask related theories, a behavioral explanation for monthly contrarian profits would be valuable. Lo and MacKinlay [1990] demonstrated that contrarian profits do not imply over reaction by presenting an alternative inefficiency that can create contrarian profits. They show that if some stocks react to common factors with a lag, then even if there is no over reaction there would be contrarian profits. However, Jegadeesh and Titman [1995a] show that the contribution of over reaction to contrarian profits is much stronger than the contribution of the alternative lagged reaction inefficiency proposed by Lo and MacKinlay [1990].

Our paper proposes a behavioral explanation for the monthly over reaction that leads to significant monthly contrarian profits. The contribution of our paper is three folds. Our first contribution is to adapt behavioral theories that are created to explain 3-12 month momentum and 3-5 year reversion to explain monthly over reaction.\(^5\)

Our second contribution is to use momentum as a test variable. Based on standard terminology in the literature we will describe a stock as high (low) momentum stock at a particular month, if during the previous 6 months leading up to that month, the stock provided much higher (lower) returns compared to the market average. One can consider momentum of a stock as a proxy for over confidence of investors about private signals\(^6\) related to the stock. Therefore analyzing monthly serial correlations of stocks with different momentums will allow us to test DHS. One can also consider low and high momentum stocks as stocks which have been receiving a stream of good news or bad news which would help us test BSV.

---

\(^5\)Chan [2003] also employs same theories to describe short term phenomenon. However, he does not provide an explanation for monthly negative serial correlation.

\(^6\)Given that market is overall net long a stock, on average participants will be more confident about their information sources after a stock performs very well.
Our third contribution is a direct implication of affirmation of DHS theory which results in low momentum stocks having a very strong negative monthly serial correlation and high momentum stocks not having any significant monthly serial correlation. To test the theories we use the news dataset collected by Chan [2003] which allows us to filter out the situations in which the returns are not public news driven. DHS claims that public news will be under reacted by confident investors; therefore if self confidence is an increasing function of momentum, we should see an under reaction (over reaction) to public news for high (low) momentum stocks. We show that this is the case. These results when combined with reversion of non-news driven returns as found in Chan [2003] yields the basic finding of our paper: Low momentum stocks are the only stocks that show monthly negative serial correlation. Moreover, we show that unlike claimed by Avramov et al. [2006], monthly negative serial correlation is not illiquidity driven. On the practical side, this study proposes a trading strategy that utilizes the fact that low momentum stocks show a very strong negative serial correlation. Returns from monthly contrarian strategy are on average 1.76% per month from 1980 to 2006.

Section 1.2 proceeds with the presentation of data and preliminary results. Section 1.3 presents our main result which is strong negative monthly serial correlation of low momentum stocks and insignificant monthly serial correlation of high momentum stocks. Section 1.4 disaggregates the results of previous section based on public news driven and non-news driven returns and presents the distinction. Section 1.5 analyzes the different response of high momentum and low momentum stocks to public news and implications for behavioral theories. Section 1.6 checks for robustness including controlling for the effects of illiquidity, turnover and systemic risk. Section 1.7 concludes.
1.2 Data and Preliminary Results

1.2.1 Data Description

Summary Statistics of the dataset can be found in table 1.1. All of the variables except the Number of News is obtained from WRDS (Wharton Research Data Services). In order to avoid survivorship bias we use monthly de-listed returns.

Our proxy for the illiquidity is the Amihud [2002] measure, which is computed as the absolute price change per dollar of daily trading volume. Monthly turnover is defined as the dollar trading volume divided by market capitalization as in Avramov et al. [2006]. Following Jegadeesh and Titman [2001], we exclude stocks with a share price below $5 at the portfolio formation date to make sure that the results are not driven by small, illiquid stocks or by the bid-ask bounce. Following Jegadeesh and Titman [2001], we exclude stocks with a share price below $5 at the portfolio formation date to make sure that the results are not driven by small, illiquid stocks or by the bid-ask bounce. Our momentum measure for a stock for month t, is the compounded raw returns for the past six months (from month t-7 to month t-1) for that stock.

An important part of our analysis depends on the differentiation between returns that can be associated with public news and returns that seem to be realized without any important news announcement about the company. To make this differentiation, we use Chan [2003] news dataset. This dataset covers a random sample of approximately one-quarter of all CRSP stocks. Chan [2003] tracked news about these stocks from Dow Jones Interactive Publications Library of past newspapers looking at only publications with over 500,000 current subscribers. For each stock covered, the dataset collects the dates at which the stock was mentioned in the headline or lead paragraph of an article of one of the publications covered.

News dataset from Chan [2003] covers 1980-2000 period. All of the other variables are from 1980 to 2006. Minimum values of company size and turnover turn out to

---

7Since dataset from Chan [2003] covers a random sample of approximately one-quarter of all CRSP stocks, in parts of the analysis where we use this dataset, we did not use any price cutoffs to avoid insufficient observation problems. The results are robust to different price cutoffs.

8Our results are similar if we use 3, 6 or 12 months to measure momentum. The results for using alternative definitions of momentum are available upon request.

9We thank Wesley S. Chan for sharing his dataset.

10For a more detailed description of the dataset see Chan [2003]
be very close to 0 and this might raise concern about the implementability of the strategies we propose in this paper. However, in our analysis we show that our results are not restricted to particular size and turnover quantiles.

1.2.2 Preliminary Results

Our preliminary analysis documents the existence of two widely reported market imperfections in our dataset. One of these is the negative first-order serial correlation (reversal) reported by Jegadeesh [1990]. The other is the momentum effect documented by Jegadeesh and Titman [1993] which implies positive effect of previous six month performance on future returns. Before proceeding to the combination of these two main factors it will be better to see the existence of reversal and momentum independently in our dataset. Table 1.2 presents portfolio strategies that utilize momentum and monthly reversal anomalies. Portfolio strategies consist of sorting stocks into equally weighted decile portfolios based on momentum and monthly stock return during month t. The reported numbers are time series averages of the performance of these portfolios during month t+1.

The results at 1.2 clearly indicate that consistent with Jegadeesh [1990] and Jegadeesh and Titman [1993] we find a significant negative serial correlation of monthly returns and momentum. Best performing stocks during month t (P10) under perform worst performers of the the month (P1) by 2.5% with a t-statistic 7.21 during month t+1. Also, stocks that has been performing very well during the past six month tend to continue to over perform stocks that has not been performing well by 1% with t-statistic of 2.99.

1.3 Monthly Reversion of

High and Low Momentum Stocks

Jegadeesh [1990] suggested that negative serial correlation in monthly stock returns is a general phenomenon observed across the entire cross-section of stocks. Jegadeesh
[1990] analyzed serial correlation for different size portfolios, however he did not report any significant difference in the serial correlation coefficient across different sized stocks other than for the month of January.\textsuperscript{11}

One of the main findings of our study is that monthly serial correlation is not observed across the entire cross-section of stocks. Low momentum stocks show significant negative serial correlation whereas high momentum stocks do not show any significant correlation. In order to document the asymmetry between high momentum and low momentum stocks we use Fama-MacBeth Regression and portfolio approaches.

1.3.1 Fama-MacBeth Regressions

In order to see the effect of momentum on monthly reversion, for every month $t$, we sort stocks into quintiles based on their momentum. Stocks which are at the bottom (top) quintile of the market universe based on their momentum are ranked M1 (M5). Then for stocks within a particular momentum ranking, for every month, we run cross-sectional regressions:

$$R_{t+1} = \alpha_t + \beta_{1t} R_t + \beta_{2t} SIZE_t + \beta_{3t} BM_t + \epsilon_t$$  \hspace{1cm} (1)

Where $R_{t+1}$ is the dividend included return for month $t+1$, $R_t$ is the dividend included return for month $t$, $SIZE_t$ is the log of market capitalization at month $t$ and $BM$ is the log of book to market value at month $t$ (book value\textsuperscript{12} of equity divided by market value of equity).

We estimate parameters with monthly cross-sectional regressions for stocks that are in a particular momentum ranking and use the time series of the estimates to

\textsuperscript{11}For January, Jegadeesh [1990] reports that absolute value of the serial correlation coefficient for small firms is bigger than the larger firms.

\textsuperscript{12}Book value is computed as in Fama and French [2002] and measured at the most recent fiscal year-end that precedes the calculation date of market value by at least three months. Book value is defined as total assets (data 6) minus total liabilities (data 181) plus balance sheet deferred taxes and investment tax credit (data 35) minus the book value of preferred stock. Depending on data availability, the book value of preferred stock is based on liquidating value (data 10), redemption value (data 56), or carrying value (data 130), in order of preferences. We exclude firms with negative book values.
calculate a final parameter estimate and t-statistic as in Fama and MacBeth [1973]. The coefficients and t-statistics (in parentheses) are presented in Table 1.3. The striking difference between the monthly serial correlation of low momentum (M1) and high momentum (M5) stocks reveals itself in these regressions.

Monthly serial correlation for low (high) momentum stocks is -0.049 (-0.003) with a t-statistic -9.93 (-0.53). We observe from Panel B that, this difference between low and high momentum stocks is robust to controlling for size and book to market effects. Moreover a declining pattern in the magnitude of the serial correlation is observable as one goes from low momentum to high momentum stocks. The conclusion is that while we observe a strong negative serial correlation for low momentum stocks, this negative correlation disappears as momentum increases.

1.3.2 Portfolio Returns

To understand the economic significance of these results we also analyzed the returns of a related portfolio strategy. Following Jegadeesh and Titman [1993] we assign stocks to portfolios based on certain characteristics such as momentum and monthly performance. This is a standard approach in finance which is used to reduce the idiosyncratic noise in our results. At the end of each month t, we sort stocks into quintiles based on their momentum where M1 denotes the bottom quintile and M5 denotes the top quintile. Independently, we also sort stocks based on their total (dividends included) returns during month t. Stocks which are at the bottom (top) quintile based on their month t returns are ranked R1 (R5). This independent sort implies creation of 25 equally weighted portfolios every month. Then we look at the time series averages of the month t+1 returns for these portfolios.

Table 1.4 summarizes our findings about the effect of momentum on monthly negative serial correlation. The results from these portfolio returns confirm our findings from Fama-MacBeth regressions. Looking at Panel A, one can observe that within low momentum stocks there is a significant return predictability based on the negative serial correlation of returns. A monthly contrarian strategy that specializes in low momentum (M1) stocks and buys the monthly losers (R1) and sells the monthly
winners (R5) of low momentum stocks has provided monthly returns of 2.06\% during 1980-2006 with a t-value of 7.88.\textsuperscript{13} Whereas the same strategy, when applied to high momentum stocks only provided an insignificant 0.06\% monthly return during the same period.

One could argue that the difference among the high momentum and low momentum stocks might be due to difference in returns at month t. If low momentum makes the stocks more volatile than the monthly losers and winners for low momentum stocks are going to have more extreme returns compared to their high momentum counterparts. In that case, it is possible to expect a higher predictability for low momentum stocks compared to high momentum stocks even though monthly serial correlation is uniform across momentum quintiles. However, the results in Table 1.3 do not support this view as low momentum stocks have obviously stronger negative serial correlation coefficient. Moreover, in Panel B of Table 1.4 we look at the difference between mean month t returns of monthly performance quintiles for low and high momentum stocks to see if low momentum stocks show more extremities during month t. Low monthly performers of low momentum stocks on average perform 0.76\% less than low monthly performers of high momentum stocks with a t-statistic of 2.3. High monthly performers of low momentum stocks on average perform 0.95\% more than their high momentum counterparts, however this result is not significant. Even though the results imply that low momentum stocks show more extreme movements, the difference compared to high momentum stocks is simply not large enough to explain the different behavior of high and low momentum stocks during month t+1.

\textbf{1.4 News and No News}

In the previous section, we have documented that monthly returns of low momentum stocks have negative serial correlation whereas we do not observe any significant predictability for monthly returns of high momentum stocks. In this section, we will
document that a significant driver of this result is the different response of high and low momentum stocks to public news announcements.

To put the results in this section in context, it is important to mention the relevant findings of Chan [2003]. To the best of our knowledge Chan [2003] is the first paper to document the importance of public news for monthly reversion in a large dataset. The focus in Chan [2003] is longer horizons and differently from our study, the role of momentum in monthly reversion is not investigated. However, Chan [2003] documents that whether a stock has received public news or not during month $t$ is a key factor in determining the reversion experienced by the stock during the following month.

For every month $t$, Chan [2003] separates his data into two groups: Stocks that were mentioned in the headline or lead paragraph of an article from a publication with more than 500,000 current subscribers during month $t$ (news stocks) and the remaining stocks (no news stocks). He then looks at monthly serial correlation patterns for these two different groups. Conclusion is that monthly negative serial correlation is significant for only no news stocks.

In order to see the effect of momentum for news and no news groups, we utilized the news dataset collected by Chan [2003]. To analyze the different response of high momentum and low momentum stocks to public announcements we will conduct an analysis which is very similar to that of the previous section. As in Chan [2003] we will divide stocks into two groups: Stocks that were mentioned in the headlines (news stocks) based on Chan’s dataset and stocks that were not mentioned in the headlines. After this division, we will repeat the portfolio analysis of subsection 1.3.2 with news and no news stocks. Every month we sort stocks into quintiles according to their momentum for both news and no news groups. Independently we also sort both news and no news stocks based on their monthly performance into three quantiles.

\[Pritamani and Singal [2001] collected daily news stories from the Wall Street Journal and Dow Jones News Wire for a subset of stocks from 1990 to 1992 where they used strict filters for trading volume, volatility, size, and price that resulted in a subset of about 1% of the NYSE/AMEX universe.\]

\[Average monthly return for a monthly contrarian strategy is 0.33\% with a t-statistic 1.34 for news stocks compared to 1.83\% with a t-statistic 6.92 for no news stocks.\]

\[Since in this analysis, we only have a subset of stocks in CRSP database, we are sorting stocks into three quantiles instead of quintiles in order to have adequate number of stocks in our portfolios. However the results are robust to sorting into quintiles or quartiles.\]
By looking at the impact of news for low momentum and high momentum stocks in table 1.5, we can tell a more complete story about monthly mean reversion. Main finding in Chan [2003] holds here for low momentum (M1) stocks, where monthly serial correlation strengthens when there is no news. According to table 1.5 Panel A and Panel B, a contrarian strategy of going long R1 and short R3 provides a monthly return of 1.5% for news stocks with low momentum whereas the same strategy provides 1.9% for no news stocks with low momentum. We have a more different result for high momentum stocks in which we actually see positive serial correlation for news stocks. A contrarian strategy of going long R1 and short R3 provides a monthly return of -0.5% for news stocks with high momentum whereas the same strategy provides 0.6% for no news stocks with high momentum. All of the returns are significant at 10% level.

Since Chan [2003] does not differentiate between high momentum and low momentum stocks, he reports an overall under reaction to news. In table 1.5 Panel B, different from Chan [2003] results, we actually show that the monthly serial correlation of news stocks depend on these stocks’ previous performance. In particular, low momentum stocks that were mentioned in the media in a given month show negative serial correlation, whereas high momentum stocks that were mentioned in the media show positive serial correlation. These results imply over reaction to news for low momentum stocks and under reaction to news for high momentum stocks. These results have implications for the existing theories of over reaction and under reaction to public news. We will go into a more detailed discussion of these results and their implications in the next section.

Another interesting observation from table 1.5 is the difference in magnitude in monthly serial correlation for high momentum and low momentum no news stocks. One can clearly see that among no news stocks as momentum increases, the profit of the contrarian strategy significantly decreases and low momentum stocks have a much stronger negative serial correlation compared to high momentum stocks. In particular, the contrarian strategy yields 1.9% (0.6%) per month with a t statistic of 5.92 (1.71) for low (high) momentum stocks.
Looking at the results in table 1.5, one can have a better understanding of our findings in section 1.3 where we did not distinguish stocks by public news. When there is no news, monthly returns show a stronger negative serial correlation for low momentum stocks compared to high momentum stocks. Moreover, when there is news; low momentum stocks continue to show negative serial correlation whereas high momentum stocks show a positive serial correlation. Therefore, when we look at the returns without differentiating with respect to the existence of news; positive correlation for high momentum news stocks neutralizes the negative correlation for high momentum no news stocks and we do not observe any significant predictability for high momentum stocks.

1.5 Response to News: 
Overreaction of Low Momentum and 
Under reaction of High Momentum Stocks

Among the behavioral models of DHS and BSV; only the setup of DHS can explain the above reported discrepancy between low momentum and high momentum stocks. One of the main goals of DHS is to explain under reaction to news phenomenon. Since our findings suggest that only high momentum stocks under react to news, original DHS model cannot explain the findings of our paper. However, DHS model is built in a way that under reaction to news is a function of how confident the investors are about their private information sources. If investors are very confident about their private information, they underestimate the importance of public news announcement and hence under react to it. In their model, DHS never consider the flip side of this argument. That is, if investors do not have enough confidence in their own private information they would overestimate the importance of a news announcement and would overreach to it. Therefore, if investors believe that their private signals are more precise than they actually are, they will under react to public information. Whereas if investors believe that their private signals are less precise than they actually are, they
will over react to public information. Given that investors are on average net long any particular stock, it is reasonable to assume that after a reasonably long period of very bad performance (low momentum), investors will have less confidence to their own private information sources about this stock and will start to put too much weight in public news announcements which will result in over reaction. Whereas after a reasonably long period of very good performance (high momentum), investors will be overly confident about their own private information sources and will start to underestimate the significance of public news announcements which will result in under reaction. This is exactly the pattern we observe in our analysis.

DHS also mentions that stock performance can affect the investor confidence. However, they do this in the context of including biased self attribution in their model. They model biased self attribution as follows: The confidence of the investors grows if their private signals prove to be accurate, however it does not fall commensurately when their private signals prove to be wrong. In our context, this would imply the following scenario: After high momentum, there is a lot of confidence in private information resources and hence there is under reaction to public news. On the other hand, after low momentum, the confidence of the investors does not decline as much compared to the increase in the confidence of high momentum investors, therefore we should not see any significant over reaction for low momentum stocks. Given that in our analysis we find both significant over reaction for low momentum stocks and significant under reaction for high momentum stocks, we conclude that our findings do not support the biased self attribution aspect of DHS model.

The results we present has also relevance for the theory presented by BSV. BSV proposes a model in which the over reaction or under reaction to news depends on the past performance of the stock, hence our setup enables a direct test of their model. Their basic idea is based on the assumption that investors do not know the statistical properties of firms’ earnings. In their model, firms’ earnings follow a random walk. However, investors believe that firms’ earnings are either positively or negatively correlated even though they are not sure which one of these is the correct statistical model. Investors use Bayesian learning to infer about the statistical
properties of firm's earnings. Therefore, after a string of bad news or good news, investors infer that firm's earnings are positively correlated. This means that when there is a positive or negative shock during an earnings announcement, investors will overestimate the impact of earnings for future firm value as they think that there is a positive correlation between this quarters earnings and next quarters earnings. This will cause an over reaction to the earnings announcement. However, if lately stock received conflicting news and earnings surprises, investors will infer that firm's consecutive earnings are negatively correlated. This will create an under reaction to the earnings announcement.

Therefore, the basic implication of BSV model for our case is as follows. After high momentum (M5) and low momentum (M1) we should see an over reaction to news, whereas if there is no significant momentum (M3), we should see an under reaction. However, the results we present in table 1.5 panel A contradict this prediction. BSV predicts that if one follows a monthly contrarian strategy for stocks that received news coverage during a particular month; one should observe that high momentum and low momentum contrarian strategy profits should be positive whereas for stocks that have no significant previous momentum, contrarian strategy profits should be negative. Instead of this, we see an almost continuous decline in the profits of the contrarian strategy. Profits of the contrarian strategy are 1.5%, 0.2%, -0.5%, 0.0% and -0.5% for M1, M2, M3, M4 and M5 respectively. This suggests a decline in the over reaction as one goes from low momentum to high momentum whereas BSV predicts that M1 and M5 contrarian profits should have been approximately equal and much higher than M3 contrarian profits (A U-shaped relationship).

1.6 Controlling for Various Cross-Sectional Effects and Robustness Checks

So far we have documented the crucial role of momentum for monthly reversion. We have also investigated the role of public news announcements in our findings. Overall
our results indicate that both firms’ past performance and the existence of news had an effect on monthly reversion. In this section, we will check the robustness of our results to controlling other factors which are documented to effect monthly reversion: illiquidity and turnover. Also, we will explore the risk based explanations of our findings.

1.6.1 Illiquidity

Avramov et al. [2006], claims that short-run contrarian profits are driven by illiquidity as in the rational equilibrium framework of Campbell et al. [1993]. Avramov et al. [2006] documents that weekly negative serial correlation is stronger for stocks with high illiquidity. They claim that illiquidity is also a driver of monthly return reversals. Since momentum and illiquidity are negatively correlated, in our results momentum might be capturing the effect of illiquidity and the results might be driven by illiquidity rather than momentum. In order to control for illiquidity, following Avramov et al. [2006] we sort stocks into quartiles independently for momentum and illiquidity. IL1 denotes the lowest illiquidity quartile which includes liquid stocks whereas IL4 denotes highly illiquid stocks. We then perform the following cross-sectional Fama-MacBeth regressions for each of the sixteen categories:

\[ R_{t+1} = \alpha_t + \beta_t R_t + \epsilon_t \]  

If our results in table 1.3 are driven by illiquidity rather than momentum, then we should not observe any significant difference among \( \beta \) across groups with same illiquidity rank. Table 1.6 presents time series averages of the \( \beta_t \) coefficients from Fama-MacBeth regressions. Panel A and Panel B confirms previous findings about the effect of momentum and illiquidity on monthly negative serial correlation. In Panel A, as momentum increases the reversal coefficient weakens from -0.045 with a t-statistic -9.76 to -0.001 with a t-statistic -0.12. In Panel B, as illiquidity increases the reversal coefficient strengthens from -0.004 with a t-statistic -0.6 to -0.029 with a t-statistic -6.65. Therefore illiquid and low momentum stocks show stronger negative serial
correlation.

However, Panel C presents a more complete story. According to the results, the effect of momentum is always significant regardless of the illiquidity level. Among liquid stocks, IL1, the average $\beta$ from equation (2) increases monotonically from $-0.025$ with a t-statistic -3.18 to 0.004 with a t-statistic 0.46. Similarly, among illiquid stocks, IL4, the average $\beta$ from equation (2) increases monotonically from $-0.056$ with a t-statistic -9.5 to -0.004 with a t-statistic -0.68. The differences are significant at 1% level. These results indicate that our findings in table refl3 are not driven by illiquidity. The effect of momentum is always significant. Illiquidity increases the effect of momentum but cannot be the whole driver of our findings. Therefore our results are robust to controlling for illiquidity.

On the other hand, our results in Panel C reveal another important fact about the effect of illiquidity on monthly negative serial correlation. While momentum has a significant impact among all illiquidity groups; illiquidity increases negative serial correlation only among low momentum stocks. Within high momentum group, M4, illiquidity has no significant effect on monthly negative serial correlation of stock returns. However, for low momentum stocks, M1, the average $\beta$ decreases monotonically from $-0.025$ with a t-statistic of -3.18 to -0.056 with a t-statistic of -9.5 with illiquidity. The difference is significant at 1% level. According to these results, Avramov et al. [2006] argument is not complete. They mainly argue that the predictability in short-horizon returns occurs because of illiquid market conditions. While we also conform their findings that illiquidity effects monthly reversion, we demonstrate that its effect is only limited to low momentum stocks. But the effect of momentum is much stronger compared to illiquidity and does not depend on conditioning on illiquidity.

1.6.2 Turnover

In their rational equilibrium framework Campbell et al. [1993] (henceforth, CGW) argue that non-informational trading causes price movements that show negative autocorrelation. In CGW framework, non-informed trading is accompanied by high trading volume, whereas informed trading is accompanied by low trading volume.
Using a sample of NASDAQ stocks, Conrad et al. [1994] (henceforth, CHN) finds support for CGW model at weekly frequency. Cooper [1999], on the other hand, uses a large sample of large NYSE-AMEX stocks and finds that weekly reversion decreases with trading activity.\footnote{Since both CHN and Cooper Cooper [1999] are dealing with weekly reversion we cannot directly compare their results with ours. Nevertheless, CGW model may still have implications for our findings as their model can also be applied to monthly periods.} Cooper [1999] results are in line with the asymmetric information model of Wang [1994] which argues that price continuations are accompanied by high trading volumes when informed investors condition their trades on private information.

In order to control for the amount of trading activity, we use monthly turnover following Avramov et al. [2006]. Turnover is defined as monthly dollar trading volume divided by market capitalization. We exclude all NASDAQ stocks to avoid problems with inflated trading volume due to double-counted dealer trades as suggested in Lee and Swaminathan [2000]. Our analysis will follow the same methodology as the previous subsection. We sort stocks into quartiles independently for momentum and turnover where T1 (T4) denotes the lowest (highest) turnover quartile. This implies sorting stocks into 16 groups every month. We then estimate the equation (2) cross sectionally for each of the 16 groups. If our results in table 1.3 are driven by turnover rather than momentum, then we should not observe any significant difference among $\beta$ across groups with same turnover rank.

Table 1.7 presents the $\beta$ coefficients from Fama-MacBeth regressions. Panel A of table 1.7 confirms findings of Avramov et al. [2006] in monthly frequency.\footnote{This result is at odds with CWG and weekly findings of Avramov et al. [2006].} We find a positive relationship between turnover and monthly reversion. In Panel B, we show that the effect of momentum is robust to controlling for turnover. According to the results the effect of momentum is always significant regardless of the turnover level. Among low turnover stocks, T1, the average $\beta$ from equation (2) increases from -0.050 with a t-statistic -6.20 to -0.032 with a t-statistic -5.19 with momentum. Similarly, among high turnover stocks, T4, the average $\beta$ increases from -0.019 with a t-statistic -1.85 to 0.017 with a t-statistic 2.76 with momentum. The differences are
significant at 1% level. Therefore our findings in table 1.3 are not driven by turnover. On the other hand, the effect of turnover is still present even if we control for the effect of momentum. This confirms the monthly result from Avramov et al. [2006] that turnover mitigates monthly reversion.

1.6.3 Risk Factor Loadings of Monthly Contrarian Strategy Profits

Our study explains the monthly contrarian strategy profits based on behavioral theories. However, an alternative explanation can be risk related. It might be the case that the return from low momentum monthly contrarian strategy is reward for the risks assumed by implementing this strategy. To test this hypothesis, we analyze the correlation of returns to the monthly contrarian strategy for each momentum quintile and nondiversifiable risk factors in the economy. Specifically we run a time series regression of returns to monthly contrarian strategy for five momentum quintiles on three Fama-French factors and momentum factor:

\[
PROFIT = \alpha + \beta_1 MKTRF + \beta_2 SMB + \beta_3 HML + \beta_4 UMD + \epsilon
\]  

(MKTRF is the value weighted return on all NYSE, AMEX and NASDAQ stocks (from CRSP) minus the one month Treasury bill rate (from Ibbotson Associates). SMB (size factor), HML (book to market factor) and UMD (momentum factor) are from Kenneth R. French’s web site.

The factor loadings are reported in table 1.8. The results reveal no significant positive factor loading for the monthly contrarian strategy. Therefore, risk based explanations of our findings are not sustainable as the profits does not seem to be positively correlated with the major risk factors considered in the literature. Among the low momentum groups, M1, the intercept from regression is 0.021 with a t-statistic

---

19 This is the long R1 short R5 strategy described in table 1.4.
20 In the empirical asset pricing literature these four factors are considered to represent major nondiversifiable risks assumed by stock market investors.
21 http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
8.38 indicating that the four factor model cannot explain away the profits. The effect of momentum is still significant even after controlling for various risk factors as the intercept term for high momentum monthly contrarian strategy is insignificant.

1.6.4 January Effect and Sub period Analysis

In order to control for January effect, we present the results in table 1.3 separately for the month of January and other months. We also present the results for three sub periods 1980-1989, 1990-1999 and 2000-2006.

Panel A in table 1.9 shows that January effect is a contributor to our results but it cannot explain away our findings. If we only use the data where January returns are the dependent variable we see that a very strong reversion occurs during the January for low momentum stocks which performed poorly during December. Autocorrelation coefficient for this time of the year is -0.111 with a t-statistic -7.72. However, even if we do not use this part of the year in our regressions we still see a very strong monthly reversal for low momentum stocks for the rest of the year. Autocorrelation coefficient for low momentum stocks during the rest of the year is -0.043 with a t-statistic -8.12.

Panel B in table 1.9 shows the declining profitability of low momentum monthly contrarian strategy. Negative serial correlation seems to be declining during the recent years from -0.050 to -0.032. However, needles to say more data is needed to see if this is an eventual disappearance of a market efficiency or just a temporary decline.

1.7 Conclusion

Most of the behavioral finance literature focuses on explaining long term predictabilities and inefficiencies whereas illiquidity and bid-ask spread related explanations is considered enough to explain weekly and monthly anomalies. Our contribution is to make the case for behavioral theories to explain monthly reversion.

We use low momentum as a proxy for investor over-reliance to public news. This implies over reaction of low momentum stocks to public news as in DHS. We document that this indeed is the case. Moreover, the concentration of monthly reversion in low
momentum stocks is robust to controlling for illiquidity, turnover and major risk factors. Therefore, we cannot reject our behavioral explanation in favor of market micro structure explanations of monthly negative serial correlation.

The concentration of monthly reversion in low momentum stocks also allows us to devise a profitable monthly contrarian strategy which focuses only on low momentum stocks. It would be interesting to see if this asymmetry between low momentum and high momentum stocks is specific to US stock market or international stock markets contain similar asymmetries.
1.8 Appendix

Table 1.1: Summary Statistics

Monthly return, Stock Price, Company Size and are from WRDS. Turnover and Illiquidity are calculated from daily return and volume data from WRDS\textsuperscript{a}. Number of News is from Chan [2003]. Stocks are included in the data if they have return data for the past six months and if they have share price higher than five dollars. All the data except Number of News is from 1980 to 2006. Number of News is from 1980 to 2000.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Observations</th>
<th>Monthly Mean</th>
<th>Monthly Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly Return</td>
<td>1,568,351</td>
<td>0.02</td>
<td>0.14</td>
<td>-0.85</td>
<td>12.61</td>
</tr>
<tr>
<td>Stock Price</td>
<td>1,568,351</td>
<td>29.23</td>
<td>643.72</td>
<td>5.00</td>
<td>109990.00</td>
</tr>
<tr>
<td>Company Size (In Billion $)</td>
<td>1,568,351</td>
<td>1.48</td>
<td>8.73</td>
<td>0.00</td>
<td>602.43</td>
</tr>
<tr>
<td>Illiquidity</td>
<td>1,443,205</td>
<td>0.66</td>
<td>3.15</td>
<td>0.00</td>
<td>774.62</td>
</tr>
<tr>
<td>Turnover</td>
<td>948,231</td>
<td>0.86</td>
<td>15.62</td>
<td>0.00</td>
<td>93.93</td>
</tr>
<tr>
<td>Number of News</td>
<td>230,805</td>
<td>1.37</td>
<td>1.48</td>
<td>0.00</td>
<td>11</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Illiquidity and turnover are calculated as in Avramov et al. [2006].
Table 1.2: Monthly Reversion and Momentum: Portfolio Approach

To create reversal portfolios P1-P10, at the end of every month t, we sort stocks into equally weighted decile portfolios based on their month t returns, then we observe the performance of these deciles during month t+1. To create momentum portfolios P1-P10, at the end of every month t, we sort stocks into equally weighted decile portfolios based on their momentum, then we observe the performance of these deciles during month t+1. Both for reversion and momentum the reported values are time series averages of month t+1 returns and P1 (P10) denotes the portfolio with stocks that have the highest (lowest) values of the sort criterion. We also report the profits from longing P1 and shorting P10. T-statistics of the profits from this long-short strategy is calculated from time series returns and is heteroskedasticity and autocorrelation robust.

<table>
<thead>
<tr>
<th>Sort Ranking</th>
<th>Reversion</th>
<th>Momentum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.0265</td>
<td>0.0079</td>
</tr>
<tr>
<td>P2</td>
<td>0.0121</td>
<td>0.0067</td>
</tr>
<tr>
<td>P3</td>
<td>0.0115</td>
<td>0.0098</td>
</tr>
<tr>
<td>P4</td>
<td>0.0121</td>
<td>0.0121</td>
</tr>
<tr>
<td>P5</td>
<td>0.0119</td>
<td>0.0121</td>
</tr>
<tr>
<td>P6</td>
<td>0.0119</td>
<td>0.0126</td>
</tr>
<tr>
<td>P7</td>
<td>0.0122</td>
<td>0.0122</td>
</tr>
<tr>
<td>P8</td>
<td>0.0118</td>
<td>0.0131</td>
</tr>
<tr>
<td>P9</td>
<td>0.0096</td>
<td>0.0157</td>
</tr>
<tr>
<td>P10</td>
<td>0.0013</td>
<td>0.0183</td>
</tr>
<tr>
<td>P1-P10</td>
<td>0.0252</td>
<td>-0.0104</td>
</tr>
<tr>
<td>t-value</td>
<td>(7.21)</td>
<td>(-2.99)</td>
</tr>
</tbody>
</table>
Table 1.3: Interaction of Momentum and Monthly Reversion: Fama-MacBeth Regressions

At the end of each month $t$, stocks which are at the bottom (top) quintile of the market based on their momentum are ranked M1 (M5). At Panel A, every month for each momentum rank, return at month $t+1$ is cross sectionally regressed on a constant and return at month $t$. At Panel B, control variables SIZE and BM are included in the regression. SIZE is the log of the market capitalization at month $t$, BM is the log of book to market value which is calculated as in Fama and French [2002]. The parameter estimates and heteroskedasticity and autocorrelation robust $t$-statistics are obtained from time series of the corresponding regression coefficients as in Fama and MacBeth [1973]. $R^2$ is the monthly average of the adjusted $R^2$ from the cross-sectional regressions.

<table>
<thead>
<tr>
<th>Momentum Rank</th>
<th>intercept</th>
<th>$\text{RET}_t$</th>
<th>SIZE</th>
<th>BM</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.007</td>
<td>-0.049</td>
<td></td>
<td></td>
<td>1.15%</td>
</tr>
<tr>
<td></td>
<td>(1.70)</td>
<td>(-9.93)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.012</td>
<td>-0.013</td>
<td></td>
<td></td>
<td>0.94%</td>
</tr>
<tr>
<td></td>
<td>(4.77)</td>
<td>(-2.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.013</td>
<td>0.002</td>
<td></td>
<td></td>
<td>1.02%</td>
</tr>
<tr>
<td></td>
<td>(5.45)</td>
<td>(0.36)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.013</td>
<td>0.005</td>
<td></td>
<td></td>
<td>0.92%</td>
</tr>
<tr>
<td></td>
<td>(5.16)</td>
<td>(0.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>0.017</td>
<td>-0.003</td>
<td></td>
<td></td>
<td>0.87%</td>
</tr>
<tr>
<td></td>
<td>(4.86)</td>
<td>(-0.53)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Momentum Rank</th>
<th>intercept</th>
<th>$\text{RET}_t$</th>
<th>SIZE</th>
<th>BM</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.012</td>
<td>-0.051</td>
<td>-0.001</td>
<td>0.003</td>
<td>3.05%</td>
</tr>
<tr>
<td></td>
<td>(1.15)</td>
<td>(-9.44)</td>
<td>(-1.25)</td>
<td>(3.06)</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.022</td>
<td>-0.013</td>
<td>-0.001</td>
<td>0.002</td>
<td>3.05%</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(-2.28)</td>
<td>(-1.58)</td>
<td>(2.84)</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.028</td>
<td>-0.003</td>
<td>-0.001</td>
<td>0.001</td>
<td>3.49%</td>
</tr>
<tr>
<td></td>
<td>(3.28)</td>
<td>(-0.41)</td>
<td>(-1.84)</td>
<td>(1.87)</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.043</td>
<td>0.005</td>
<td>-0.002</td>
<td>0.001</td>
<td>3.27%</td>
</tr>
<tr>
<td></td>
<td>(4.35)</td>
<td>(0.88)</td>
<td>(-2.79)</td>
<td>(1.60)</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>0.058</td>
<td>-0.004</td>
<td>-0.002</td>
<td>0.001</td>
<td>2.9%</td>
</tr>
<tr>
<td></td>
<td>(4.79)</td>
<td>(-0.68)</td>
<td>(-2.65)</td>
<td>(0.67)</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.4: Interaction of Momentum and Monthly Reversion: Portfolio Approach

At the end of each month $t$, we sort stocks into quintiles based on their momentum where $M1$ denotes the bottom quintile and $M5$ denotes the top quintile. Independently, we also sort stocks based on their total (dividends included) returns during month $t$. Stocks which are at the bottom (top) quintile based on their month $t$ return are ranked $R1$ ($R5$). This independent sort implies creation of 25 equally weighted portfolios every month. The reported results in Panel A are the time series averages of the month $t+1$ portfolio returns for these portfolios. We also report the results of going long $R1$ and short $R5$ for each momentum quintile. In Panel B, we report the difference between the average month $t$ returns of high and low momentum stocks keeping monthly performance quintile constant. Reported $t$-statistics are heteroskedasticity and autocorrelation robust.

<table>
<thead>
<tr>
<th>PANEL A</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.0122</td>
<td>0.0145</td>
<td>0.0129</td>
<td>0.0128</td>
<td>0.0191</td>
</tr>
<tr>
<td>R2</td>
<td>0.0090</td>
<td>0.0137</td>
<td>0.0133</td>
<td>0.0129</td>
<td>0.0174</td>
</tr>
<tr>
<td>R3</td>
<td>0.0073</td>
<td>0.0120</td>
<td>0.0123</td>
<td>0.0139</td>
<td>0.0165</td>
</tr>
<tr>
<td>R4</td>
<td>0.0026</td>
<td>0.0113</td>
<td>0.0124</td>
<td>0.0134</td>
<td>0.0164</td>
</tr>
<tr>
<td>R5</td>
<td>-0.0084</td>
<td>0.0085</td>
<td>0.0111</td>
<td>0.0132</td>
<td>0.0185</td>
</tr>
<tr>
<td>R1-R5</td>
<td>0.0206</td>
<td>0.0060</td>
<td>0.0018</td>
<td>-0.0004</td>
<td>0.0006</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(7.88)</td>
<td>(3.39)</td>
<td>(1.12)</td>
<td>(-0.21)</td>
<td>(0.24)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PANEL B</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1-M5</td>
<td>-0.0076</td>
<td>-0.0009</td>
<td>0.0034</td>
<td>0.0060</td>
<td>0.0095</td>
</tr>
<tr>
<td>$t$-value</td>
<td>(-2.3)</td>
<td>(-0.33)</td>
<td>(1.23)</td>
<td>(1.58)</td>
<td>(1.36)</td>
</tr>
</tbody>
</table>

*For robustness, we also examine stock returns based on value-weighted and log-value-weighted portfolios. The results are similar.
Table 1.5: Interaction of Momentum and Monthly Reversion for News and No News Stocks

For Panel A, we only take stocks that are in Chan’s randomized sample and that were mentioned in the headlines during month \( t \). At the end of each month \( t \), we sort these stocks into quintiles based on their momentum where \( M_1 \) denotes the bottom quintile and \( M_5 \) denotes the top quintile. Independently, we also sort stocks based on their total (dividends included) returns during month \( t \). Stocks which are at the bottom (top) \( 1/3 \) quantile based on their month \( t \) return are ranked \( R_1 \) (\( R_3 \)). This independent sort implies creation of 15 equally weighted portfolios every month. The reported results in Panel A are the time series averages of the month \( t+1 \) portfolio returns for these portfolios. We also report the results of going long \( R_1 \) and short \( R_3 \) for each momentum quintile. In Panel B, we do the identical analysis, this time for stocks that were in Chan’s randomized sample but were not mentioned in the headlines during month \( t \). Reported \( t \)-statistics are heteroskedasticity and autocorrelation robust and calculated from time series returns of long-short portfolios.

### PANEL A: News Stocks

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.0105</td>
<td>0.0128</td>
<td>0.0100</td>
<td>0.0117</td>
<td>0.0166</td>
</tr>
<tr>
<td>R2</td>
<td>0.0073</td>
<td>0.0128</td>
<td>0.0131</td>
<td>0.0129</td>
<td>0.0190</td>
</tr>
<tr>
<td>R3</td>
<td>-0.0042</td>
<td>0.0105</td>
<td>0.0151</td>
<td>0.0120</td>
<td>0.0219</td>
</tr>
<tr>
<td>R1-R3</td>
<td>0.0148</td>
<td>0.0023</td>
<td>-0.0051</td>
<td>-0.0003</td>
<td>-0.0053</td>
</tr>
<tr>
<td>t-value</td>
<td>(4.64)</td>
<td>(0.94)</td>
<td>(-2.37)</td>
<td>(-0.12)</td>
<td>(-1.66)</td>
</tr>
</tbody>
</table>

### PANEL B: No News Stocks

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.0141</td>
<td>0.0151</td>
<td>0.0122</td>
<td>0.0175</td>
<td>0.0230</td>
</tr>
<tr>
<td>R2</td>
<td>0.0051</td>
<td>0.0126</td>
<td>0.0137</td>
<td>0.0154</td>
<td>0.0207</td>
</tr>
<tr>
<td>R3</td>
<td>-0.0042</td>
<td>0.0105</td>
<td>0.0151</td>
<td>0.0120</td>
<td>0.0219</td>
</tr>
<tr>
<td>R1-R3</td>
<td>0.0193</td>
<td>0.0073</td>
<td>0.0008</td>
<td>0.0040</td>
<td>0.0063</td>
</tr>
<tr>
<td>t-value</td>
<td>(5.92)</td>
<td>(2.55)</td>
<td>(0.31)</td>
<td>(1.58)</td>
<td>(1.71)</td>
</tr>
</tbody>
</table>
Table 1.6: Momentum and Illiquidity

For Panel A, every month we sort stocks into quartiles based on their momentum. For Panel B, every month we sort stocks into quartiles based on their illiquidity where IL1 denotes low illiquidity and IL4 denotes high illiquidity. For Panel C, we independently sort stocks into quartiles based on their momentum and illiquidity. Then for every one of these quartiles and for every month we run the following regression cross-sectionally:

\[ R_{t+1} = \alpha_t + \beta_t R_t + \epsilon_t \]

Time series averages of \( \beta \) coefficients are reported as in Fama and MacBeth [1973]. Reported t-statistics are heteroskedasticity and autocorrelation robust.

<table>
<thead>
<tr>
<th>PANEL A: Momentum</th>
<th>PANEL B: Illiquidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>IL1</td>
</tr>
<tr>
<td>-0.045</td>
<td>-0.004</td>
</tr>
<tr>
<td>(-9.76)</td>
<td>(-0.6)</td>
</tr>
<tr>
<td>M2</td>
<td>IL2</td>
</tr>
<tr>
<td>-0.007</td>
<td>-0.012</td>
</tr>
<tr>
<td>(-1.21)</td>
<td>(-2.06)</td>
</tr>
<tr>
<td>M3</td>
<td>IL3</td>
</tr>
<tr>
<td>0.003</td>
<td>-0.016</td>
</tr>
<tr>
<td>(0.68)</td>
<td>(-3.31)</td>
</tr>
<tr>
<td>M4</td>
<td>IL4</td>
</tr>
<tr>
<td>-0.001</td>
<td>-0.029</td>
</tr>
<tr>
<td>(-0.12)</td>
<td>(-6.65)</td>
</tr>
</tbody>
</table>

Panel C: Momentum and Illiquidity

<table>
<thead>
<tr>
<th>IL1</th>
<th>IL2</th>
<th>IL3</th>
<th>IL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.025</td>
<td>-0.044</td>
<td>-0.045</td>
<td>-0.056</td>
</tr>
<tr>
<td>(-3.18)</td>
<td>(-6.79)</td>
<td>(-6.95)</td>
<td>(-9.5)</td>
</tr>
<tr>
<td>M2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.012</td>
<td>-0.008</td>
<td>-0.002</td>
<td>-0.015</td>
</tr>
<tr>
<td>(-1.51)</td>
<td>(-0.97)</td>
<td>(-0.28)</td>
<td>(-2.23)</td>
</tr>
<tr>
<td>M3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.007</td>
<td>0.009</td>
<td>0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>(0.99)</td>
<td>(1.14)</td>
<td>(0.9)</td>
<td>(-1.3)</td>
</tr>
<tr>
<td>M4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.004</td>
<td>-0.004</td>
<td>0.003</td>
<td>-0.004</td>
</tr>
<tr>
<td>(0.46)</td>
<td>(-0.55)</td>
<td>(0.43)</td>
<td>(-0.68)</td>
</tr>
</tbody>
</table>
Table 1.7: Momentum and Turnover

For Panel A, every month we sort stocks into quartiles based on their turnover, where T1 (T4) denotes low (high) turnover. For Panel B, we independently sort stocks into quartiles based on their momentum and turnover. Then within every one of these groups and for every month we run the following regression cross-sectionally:

\[ R_{t+1} = \alpha_t + \beta_t R_t + \epsilon_t \]

Time series averages of \( \beta \) coefficients are reported as in Fama and MacBeth [1973]. Reported t-statistics in parentheses are heteroskedasticity and autocorrelation robust.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.030</td>
<td>-0.040</td>
<td>-0.031</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(-5.44)</td>
<td>(-7.38)</td>
<td>(-5.02)</td>
<td>(-0.84)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>-0.050</td>
<td>-0.035</td>
<td>-0.029</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(-6.20)</td>
<td>(-3.54)</td>
<td>(-3.60)</td>
<td>(-1.85)</td>
</tr>
<tr>
<td>T2</td>
<td>-0.065</td>
<td>-0.044</td>
<td>-0.034</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(-7.73)</td>
<td>(-4.79)</td>
<td>(-3.80)</td>
<td>(-1.17)</td>
</tr>
<tr>
<td>T3</td>
<td>-0.051</td>
<td>-0.023</td>
<td>-0.016</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>(-5.85)</td>
<td>(-2.60)</td>
<td>(-1.90)</td>
<td>(-2.57)</td>
</tr>
<tr>
<td>T4</td>
<td>-0.032</td>
<td>-0.006</td>
<td>0.015</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(-5.19)</td>
<td>(0.85)</td>
<td>(2.22)</td>
<td>(2.76)</td>
</tr>
</tbody>
</table>
Table 1.8: Time Series Regression of Contrarian Strategy Profits on Fama-French Factors

Profit for each momentum quintile is the return to long-short portfolio calculated in Panel A of table 1.4. We run a time series regression of profits from each momentum quintile on monthly realizations of three Fama-French and momentum factors:

\[ PROFIT = \alpha + \beta_1 MKTRF + \beta_2 SMB + \beta_3 HML + \beta_4 UMD + \epsilon \]

MKTRF is the value weighted return on all NYSE, AMEX and NASDAQ stocks (from CRSP) minus the one month Treasury bill rate (from Ibbotson Associates). SMB (size factor), HML (book to market factor) and UMD (momentum factor) are from Kenneth R. French’s web site. T-statistics which are shown in parentheses are heteroskedasticity and autocorrelation robust.

<table>
<thead>
<tr>
<th>Momentum Rank</th>
<th>( \alpha )</th>
<th>market</th>
<th>size</th>
<th>book to market</th>
<th>momentum</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.021</td>
<td>-0.006</td>
<td>0.092</td>
<td>-0.289</td>
<td>-0.009</td>
<td>5.34%</td>
</tr>
<tr>
<td></td>
<td>(8.38)</td>
<td>(-0.09)</td>
<td>(0.96)</td>
<td>(-2.52)</td>
<td>(-0.15)</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.007</td>
<td>-0.000</td>
<td>0.019</td>
<td>-0.142</td>
<td>-0.010</td>
<td>2.27%</td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
<td>(0.09)</td>
<td>(0.33)</td>
<td>(-2.01)</td>
<td>(-0.25)</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.002</td>
<td>0.040</td>
<td>-0.021</td>
<td>-0.017</td>
<td>0.000</td>
<td>1.15%</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.94)</td>
<td>(-0.01)</td>
<td>(-0.87)</td>
<td>(-0.01)</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.000</td>
<td>-0.040</td>
<td>0.024</td>
<td>-0.059</td>
<td>0.019</td>
<td>0.47%</td>
</tr>
<tr>
<td></td>
<td>(-0.08)</td>
<td>(-0.73)</td>
<td>(0.3)</td>
<td>(-0.65)</td>
<td>(0.37)</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>0.001</td>
<td>-0.100</td>
<td>0.045</td>
<td>-0.063</td>
<td>0.032</td>
<td>0.95%</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(-1.47)</td>
<td>(0.48)</td>
<td>(-0.56)</td>
<td>(0.52)</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.9: January Effect and Subperiod Analysis

The analysis is identical to Panel A of table 1.3. Fama-Macbeth regressions of next months returns are run on this months returns for each momentum quintile. In Panel A, we separated data into two mutually exclusive parts. January part includes only data points where January returns are dependent variable. The rest of the data is in February-December part. In Panel B, we separated data into three mutually exclusive groups based on whether the dependent returns in regressions are realized in 1980-1989, 1990-1999 or 2000-2006. T-statistics reported in parentheses are heteroskedasticity and autocorrelation robust.

### PANEL A

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.111</td>
<td>-0.093</td>
<td>-0.076</td>
<td>-0.044</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>(-7.72)</td>
<td>(-9.77)</td>
<td>(-7.3 )</td>
<td>(-2.61)</td>
<td>(-2.24)</td>
</tr>
<tr>
<td>February-December</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.043</td>
<td>-0.006</td>
<td>0.009</td>
<td>0.010</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(-8.12)</td>
<td>(-1.13)</td>
<td>(1.56)</td>
<td>(1.69)</td>
<td>(0.34)</td>
</tr>
</tbody>
</table>

### PANEL B

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980-1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.058</td>
<td>-0.029</td>
<td>-0.012</td>
<td>-0.005</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(-6.79)</td>
<td>(-3.08)</td>
<td>(-1.23)</td>
<td>(-0.64)</td>
<td>(-2.34)</td>
</tr>
<tr>
<td>1990-1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.050</td>
<td>-0.002</td>
<td>0.019</td>
<td>0.020</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(-6.98)</td>
<td>(0.22)</td>
<td>(2.36)</td>
<td>(2.37)</td>
<td>(1.87)</td>
</tr>
<tr>
<td>2000-2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.032</td>
<td>-0.013</td>
<td>-0.004</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(-2.86)</td>
<td>(-1.23)</td>
<td>(-0.36)</td>
<td>(-0.09)</td>
<td>(-0.97)</td>
</tr>
</tbody>
</table>
Bibliography


Michael Cooper. Filter rules based on price and volume in individual security over- 

Kent Daniel, David Hirshleifer, and Avanidhar Subrahmanyam. Investor psychology 

Eugene F. Fama. Efficient capital markets: a review of theory and empirical work. 

Eugene F. Fama and Kenneth R. French. Multifactor explanations of asset pricing 

Eugene F. Fama and Kenneth R. French. Testing trade-off and pecking order predic- 

Eugene F. Fama and James D. MacBeth. Risk, return and equilibrium: empirical 

Narasimhan Jegadeesh. Evidence of predictable behavior of security returns. *The 

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling 

Narasimhan Jegadeesh and Sheridan Titman. Overreaction, delayed reaction, and 

Narasimhan Jegadeesh and Sheridan Titman. Short-horizon return reversals and the 

Narasimhan Jegadeesh and Sheridan Titman. Profitability of momentum strategies: 

Michael C. Jensen. Some anomalous evidence regarding market efficiency. *Journal of 

Charles M. C. Lee and Bhaskaran Swaminathan. Price momentum and trading vol- 

Bruce N. Lehmann. Fads, martingales, and market efficiency. *Quarterly Journal of 

Andrew W. Lo and A. Craig MacKinlay. When are contrarian profits due to stock 

Mahesh Pritamani and Vijay Singal. Return predictability following large price 

Mark Rubinstein and Paul Stephens. Rational markets: yes or no? the affirmative 


Chapter 2

Number of Instruments in GEL Estimation with Heteroskedasticity

Generalized Empirical Likelihood (GEL) estimators have theoretically attractive properties in estimating conditional moment restriction models with a large number of moment conditions. This paper answers the question of choosing the number of moment conditions in GEL estimation. The criteria is to minimize the higher order mean squared error (MSE). This can be considered an extension of Donald et al. [2002]. As an extension, heteroskedasticity is explicitly modeled and taken into account when calculating higher order MSE terms. The results of the simulation are encouraging as they demonstrate that using higher order MSE to choose number of instruments in GEL estimation reduces the dispersion of the estimator in the case of heteroskedasticity.

2.1 Introduction

As suggested by Hansen and Singleton [1982], estimating rational expectations models which are prevalent in macroeconomics and financial economics amounts to estimating parameters that satisfy a number of orthogonality conditions.

Specifically, let $p_i, (i = 1, ..., n)$, be a $1 \times m$ vector of conditioning variables or instruments observed by the agents in the economy and the econometrician at time $i$. 
Let $x_i, (i = 1, \ldots, n)$ be a $1 \times p$ vector of endogenous variables observed by the agents in the economy and the econometrician at time $i + 1$. Discrete time models of optimizing economic agents results in the following orthogonality condition:

$$E [g (z_i, \beta_0)] = 0$$

(1)

Where $\beta_0$ is a $p$ dimensional parameter vector unknown to the econometrician, expectation is taken with respect to information at time $t$, $z_i = (x_i, p_i)$ and $g : \mathbb{R}^{p+m} \rightarrow \mathbb{R}^m$ which implies that the number of orthogonality conditions is equal to the number of conditioning variables.

Both linear expectations models as in Hansen and Sargent [1982] and Hayashi and Sims [1983] as well as nonlinear expectations models as in Hansen and Singleton [1982] can be estimated by estimating equation (1). A common problem in these estimations is to choose the number of orthogonality conditions or instruments. However, dealing with this problem just by analyzing the asymptotic properties of various estimators can be misleading. This is because the tradeoffs in asymptotic properties of estimators are not the same as tradeoffs in approximate small sample properties of estimators. For example, Hansen and Sargent [1982] notes that asymptotic variance of the estimators gets smaller as one adds more instruments and since we also have consistency for any number of instruments, this provides a justification for adding as many instruments as possible. However, taking into account small sample bias and variance changes as one adds more instruments, a more accurate answer can be given to this question as pointed out by Donald and Newey [2001] and Donald et al. [2002].

To answer the question of number of instruments by taking into account the small sample tradeoffs between bias and standard error, higher order asymptotic theory as in Nagar [1959], and Rothenberg [1984] can be applied. Donald and Newey [2001] use higher order asymptotic theory to choose the number of instruments for two-stage least squares and limited information maximum likelihood estimators. Donald et al.

\footnote{For an example with a representative agent choosing consumption and investment plans to maximize expected utility in a discrete time setting see Hansen and Singleton [1982].}

\footnote{In general $\beta_0$ will depend on the utility function parameters and production constraints in the economy.}
[2002] use a similar methodology to choose the number of instruments for GMM and Generalized Empirical Likelihood (GEL, hereafter) estimators. This paper follows closely Donald and Newey [2001] and Donald et al. [2002]. Higher order mean and variance terms for GEL estimator for \( \beta_0 \) in equation (1) is derived. An approximate small sample mean squared error (MSE) is calculated using higher order mean and variance. The number of instruments are selected to minimize the MSE. Different from Donald et al. [2002] heteroskedasticity is explicitly modeled and disturbance terms are not assumed to have zero conditional third moments. Simulation results suggest that choosing optimal number of instruments based on MSE criteria improves the estimator mostly by reducing the dispersion compared to estimating \( \beta_0 \) with as many instruments as possible which extends the results in Donald et al. [2002] to problems with more general error distributions.

Section 2.2 describes the GEL estimator and explains the reasons for choosing this estimator. Section 2.3 lays out the terms in the asymptotic expansion of the GEL estimator. Section 2.4 provides the basic form of higher order mean and variance of GEL estimator. Section 2.5 describes linear model which allows simplifications in deriving higher order MSE terms. Section 2.6 describes the simulation framework and summarizes the results. Section 2.7 concludes. Section 2.8 is Appendix which is comprised of some technical assumptions, expressions for calculating higher order MSE and tables from simulation.

### 2.2 The GEL Estimator

A common approach to estimate the \( \beta_0 \) in equation (1) is two stage GMM of Hansen [1982]. Let

\[
g_i (\beta) \equiv g (z_i, \beta)
\]

\[
\hat{g} (\beta) \equiv n^{-1} \sum_{i=1}^{n} g_i (\beta)
\]

\[
\Omega \equiv E \left[ g_i (\beta_0) g_i (\beta_0)' \right]
\]


and

\[ \hat{\Omega} (\beta) \equiv n^{-1} \sum_{i=1}^{n} g_i (\beta) g_i (\beta)' \]

The first stage estimator is given by:

\[ \tilde{\beta} = \arg\min_{\beta \in B} \hat{g} (\beta)' \hat{W}^{-1} \hat{g} (\beta) \]

where \( B \) is the parameter space and \( \hat{W} \) is the random weighting matrix which obey certain regularity conditions.\(^3\) The two step GMM estimator is:

\[ \hat{\beta}_{GMM} = \arg\min_{\beta \in B} \hat{g} (\beta)' \left( \hat{\Omega} (\tilde{\beta})^{-1} \right) \hat{g} (\beta) \]

As shown in Newey and Smith [2004] GEL estimators have theoretically attractive properties compared to two stage GMM especially when a large number of moment conditions are used as in the case of Hansen and Singleton [1982] and Holtz-Eakin et al. [1988]. Asymptotic bias of GEL estimators does not grow with the number of instruments, whereas the same cannot be said for GMM. Moreover, bias corrected GEL estimators are higher order relative efficient compared to other bias corrected estimators. The GEL estimator, \( \hat{\beta}_{GEL} \), is defined as follows:

\[ \hat{\beta}_{GEL} = \arg\min_{\beta \in B} \max_{\lambda \in \Lambda_n (\beta)} \sum_{i=1}^{n} s (\lambda g_i (\beta)) \]

where, \( \Lambda_n (\beta) = \{ \lambda : \lambda' g_i (\beta) \in V, i = 1, ..., n \} \), \( s (v) \) is a concave function with domain that is an open interval \( V \) containing 0. Let \( s_j (v) = \partial^j s (v) / \partial v^j \) and \( s_j = s_j (0) \). As suggested by Newey and Smith [2004] we can assume \( s_1 = s_2 = -1 \) without loss of generality. When \( s (v) = \ln (1 - v) \), \( \hat{\beta}_{GEL} \) is equal to empirical likelihood (EL) estimator in Qin and Lawless [1994] and Owen [1988]. When \( s (v) = -\exp (v) \), \( \hat{\beta}_{GEL} \) is equal to the exponential tilting (ET) estimator in Imbens et al. [1998] and Kitamura and Stutzer [1997]. When, \( s (v) = - (1 + v)^2 / 2 \), \( \hat{\beta}_{GEL} \) is equal to the continuous updating estimator (CUE) in Hansen et al. [1996]. The original formulation of CUE

\(^3\)See Newey and Smith [2004] for details.
estimator is closely related to two stage GMM. Rather than assuming a constant weighting matrix in the second stage, CUE updates \( \hat{\Omega}(\beta) \) continuously during the search for minimum:

\[
\hat{\beta}_{\text{CUE}} = \arg \min_{\beta \in B} \hat{g}(\beta)' \hat{\Omega}(\beta)^{-} \hat{g}(\beta)
\]

where \( A^{-} \) denotes a generalized inverse of matrix \( A \), satisfying \( AA^{-}A = A \). The equivalence of the original formulation and the GEL formulation of CUE is due to Newey and Smith [2004]. A disadvantage of CUE is its lack of moments which results in large dispersion. Hausman et al. [2007] modifies CUE resulting in a less dispersed estimator in a general setup with heteroskedasticity and autocorrelation.

For \( G_i(\beta) = \partial g_i(\beta) / \partial \beta \), let

\[
m(z_i, \theta) = s_i(\lambda' g_i(\beta)) \begin{pmatrix} G_i(\beta)' \lambda \\ g_i(\beta) \end{pmatrix}, \theta = (\beta', \lambda')'
\]

Then, the first order conditions for the GEL estimation problem (2) can be written as:

\[
\sum_i m(z_i, \hat{\theta}_{\text{GEL}}) / n = 0
\]

### 2.3 Asymptotic Expansion of the GEL Estimator

In order to calculate the approximate MSE of the GEL estimator, we need to derive the leading terms of the stochastic expansion of the \( \hat{\theta}_{\text{GEL}} \). Specifically, we need to derive \( \hat{U} = O_p(n^{-1/2}), \hat{Y} = O_p(n^{-1}) \) and \( \hat{Z} = O_p(n^{-3/2}) \) where

\[
\hat{\theta}_{\text{GEL}} = \theta_0 + \hat{U} + \hat{Y} + \hat{Z} + O_p(n^{-2})
\]

This would allow us to calculate the higher order MSE, as described e.g. in Rothenberg [1984]. To express \( \hat{U}, \hat{Y} \) and \( \hat{Z} \) we will need additional notation:

\[
M_i = \partial m(z_i, \theta_0) / \partial \theta, \quad \hat{M} = \sum_i M_i / n, \quad M = E[M_i]
\]
\[ M'_i = \partial^2 m(z_i, \theta_0) / \partial \theta_r \partial \theta, \quad \hat{M}' = \sum_i M'_i / n, \quad M' = E[M'_i] \quad (4) \]

\[ M'^s = \partial^3 m(z_i, \theta_0) / \partial \theta_r \partial \theta_s \partial \theta, \quad \hat{M}^{rs} = \sum_i M'^s / n, \quad M'^s = E[M'^s] \]

\[ \hat{U}, \hat{Y}, \text{and} \hat{Z} \text{are already derived in Newey and Smith [2004] Theorem 3.4:} \]

\[ \hat{U} = \sum_{i=1}^n U_i / n, \quad U_i = -M^{-1}m(z_i, \theta_0) \]

\[ \hat{Y} = -M^{-1} \left[ (\hat{M} - M) \hat{U} + \sum_{r=1}^k \hat{U}_r M'^r \hat{U} / 2 \right] \]

\[ \hat{Z} = -M^{-1} \left[ (\hat{M} - M) \hat{Y} + \sum_{r=1}^k \left( \hat{U}_r M'^r \hat{Y} + \hat{Y}_r M'^r \hat{U} \right) / 2 \right. \]

\[ \left. + \sum_{r=1}^k \hat{U}_r \left( \hat{M}' - M' \right) \hat{U} / 2 + \sum_{r=1}^k \sum_{s=1}^k \hat{U}_r \hat{U}_s M'^s \hat{U} / 6 \right] \]

Where \( k = m + p. \)

Newey and Smith [2004] need Assumptions 1-3 in the Subsection 2.8.1 for obtaining these stochastic expansion terms.

### 2.4 Higher Order MSE of the GEL Estimator:

**Basic Form**

\[ MSE(\hat{\theta}_{GEL}) = \left( BIAS(\hat{\theta}_{GEL}) \right)^2 + VAR(\hat{\theta}_{GEL}) \]

Therefore, to calculate approximate MSE which converges to the true MSE with \( O_p(n^{-2}) \), we will need to calculate the approximate bias which converges to true bias at most as fast as \( O_p(n^{-1}) \). Without loss of generality we can assume that \( \beta_0 = 0. \)
\( \lambda_0 = 0 \), as a general property of GEL estimators.\(^4\) Therefore:

\[
BIAS\left( \hat{\Theta}_{GEL} \right) = E\left( \hat{U} \right) + E\left( \hat{Y} \right) + E\left( \hat{Z} \right) + E\left( O_p\left( n^{-2} \right) \right)
\]

Since \( \hat{Z} \) and \( O_p\left( n^{-2} \right) \), converges to zero faster than \( n^{-1} \) we can ignore these two terms. Trivially, \( E\left[ \hat{U} \right] = 0 \). Hence,

\[
BIAS\left( \hat{\Theta}_{GEL} \right) = -M^{-1} \left[ E\left[ \hat{M}\hat{U} \right] + \sum_{r=1}^{k} M' E\left[ \hat{U}_r \hat{U} \right]/2 \right]
\]

(5)

We need to calculate the approximate variance such that it would converge to the true variance as fast as \( O_p\left( n^{-2} \right) \).

\[
VAR\left( \hat{\Theta}_{GEL} \right) = E\left[ \left( \hat{\Theta}_{GEL} - BIAS\left( \hat{\Theta}_{GEL} \right) \right) \left( \hat{\Theta}_{GEL} - BIAS\left( \hat{\Theta}_{GEL} \right) \right)' \right] = E\left[ \left( \hat{U} + \hat{Y} + \hat{Z} + O_p\left( n^{-2} \right) \right) \left( \hat{U} + \hat{Y} + \hat{Z} + O_p\left( n^{-2} \right) \right)' \right]
\]

where \( \hat{Y} = \hat{Y} - E\left( \hat{Y} \right) \). Ignoring terms that approaches zero faster than \( O_p\left( n^{-2} \right) \):

\[
VAR\left( \hat{\Theta}_{GEL} \right) = E\left[ \hat{U}\hat{U}' + \hat{Y}\hat{Y}' + \hat{Y}\hat{Y}' + \hat{U}\hat{Z}' + \hat{Z}\hat{U}' \right]
\]

\[
E\left[ \hat{U}\hat{U}' \right] = \left( \sum_{i=1}^{n} U_i \right) \left( \sum_{j=1}^{n} U_j' \right)/n^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ U_i U_j' \right]/n^2
\]

\[
E\left[ U_i U_j' \right] = 0 \text{ for } i \neq j \text{ as we assume independent observations. Let } V = E\left[ U_i U_i' \right] :
\]

\[
E\left[ \hat{U}\hat{U}' \right] = \sum_{i=1}^{n} E\left[ U_i U_i' \right]/n^2 = V/n
\]

Plugging this in the above variance formula results:

\[
VAR\left( \hat{\Theta}_{GEL} \right) = V/n + (A + B + B')/n^2
\]

(6)

\(^4\)See Newey and Smith [2004].
Where \( A = n^2 E \left( \hat{Y} \hat{Y}' \right) \), \( B = n^2 E \left( (\hat{Y} + \hat{Z}) \hat{U}' \right) \)

Therefore in order to calculate MSE, we will need \( E \left( \hat{M} \hat{U} \right) \), \( \sum_{r=1}^{k} M_r E \left( \hat{U}_r \hat{U}' \right) / 2, V, A \) and \( B \). Writing these terms in terms of expressions in (4) involves introduction of some additional notation and some additional algebra. Formulas can be found in Subsection 2.8.2.

### 2.5 Linear Model: Implementable Expressions for MSE of the GEL Estimator

If we assume that the moment conditions take the familiar linear form as in Donald and Newey [2001] and Donald et al. [2002]:

\[
g_i (\beta) = p_i (y_i - x_i' \beta)
\]

This will lead to significant simplifications when deriving implementable expressions for higher order mean and variance formulas basic form of which can be found in Section 2.4 and Subsection 2.8.2. In this special case, the following will be true:

\[
G_i (\beta) = \partial g_i (\beta) / \partial \beta = -p_i x_i'
\]

\[
m (z_i, \theta) = s_1 (\lambda' p_i (y_i - x_i' \beta)) \begin{pmatrix} -x_i p_i' \lambda \\ p_i (y_i - x_i' \beta) \end{pmatrix}
\]

\[
\frac{\partial m_i (\theta)}{\partial \theta} = s_2 (\lambda' p_i (y_i - x_i' \beta)) \begin{pmatrix} -x_i p_i' \lambda \\ p_i (y_i - x_i' \beta) \end{pmatrix} \begin{pmatrix} (-\lambda' p_i x_i', p_i' (y_i - x_i' \beta)) \\ 0 \end{pmatrix}
\]

\[
+s_1 (\lambda' p_i (y_i - x_i' \beta)) \begin{pmatrix} 0 & -x_i p_i' \\ -p_i x_i' & 0 \end{pmatrix}
\]

\[
M_i = \frac{\partial m (\theta_0)}{\partial \theta} = \begin{pmatrix} 0 & x_i p_i' \\ p_i x_i' & -p_i p_i' \varepsilon_i^2 \end{pmatrix}
\]
where $\varepsilon_i = y_i - x_i'\beta_0$.

$$M = E[M_i] = \begin{pmatrix} 0 & K' \\ K & -\Omega \end{pmatrix}$$

(7)

where $K = E[p_i'x_i']$. Derivatives of $M_i$ with respect to $\beta$ and $\lambda$ take different form. Hence, they need to be considered separately. Variables in sub vector $\beta$ is going to be indexed with $t$ and $u$ (if a second variable from sub vector $\beta$ is considered), whereas variables in sub vector $\lambda$ is going to be indexed with $v$ and $w$. After taking derivatives of $\partial m_i(\theta)/\partial \theta$ and evaluating at the true parameter values $\theta_0 = (\beta_0, \lambda_0) = (\beta_0, 0)$. One can reach the following expressions for $M_{it}^t, M_{iv}^v, M_{iu}^{tu}$ and $M_{iw}^{tv}$:

$$M_{it}^t \equiv \partial^2 m(z_i, \theta_0)/\partial \beta_t \partial \theta = \begin{pmatrix} 0 & 0 \\ 0 & 2l_{it}^1 \end{pmatrix}$$

where

$$l_{it}^1 = p_i'x_i'x_{it}$$

where $x_{it}$ denotes the $t^{th}$ element of vector $x_i$.

$$M_{iv}^v \equiv \partial^2 m(z_i, \theta_0)/\partial \lambda_v \partial \theta = \begin{pmatrix} 0 & 2l_{iv}^3 \\ 2l_{iv}^5 & l_{iv}^3 \end{pmatrix}$$

$$M_{iu}^{tu} \equiv \partial^2 m(z_i, \theta_0)/\partial \beta_u \partial \beta_t \partial \theta = \begin{pmatrix} 0 & 0 \\ 0 & -2l_{itu}^4 \end{pmatrix}$$

$$M_{iv}^{tv} \equiv \partial^2 m(z_i, \theta_0)/\partial \beta_t \partial \lambda_v \partial \theta = \begin{pmatrix} 0 & -2l_{itv}^5 \\ -2l_{itv}^7 & -3l_{itv}^6 \end{pmatrix}$$

$$M_{iw}^{vw} \equiv \partial^2 m(z_i, \theta_0)/\partial \lambda_v \partial \lambda_w \partial \theta = \begin{pmatrix} -2l_{iwv}^7 & -3l_{iwv}^8 \\ -3l_{iwv}^9 & l_{iwv}^9 \end{pmatrix}$$

where

$$l_{iv}^2 = x_i'p_i'p_{iv}\varepsilon_i, \quad l_{iv}^3 = p_i'p_{iv}s_3p_{iv}\varepsilon_i^3, \quad l_{itu}^4 = p_i'x_{it}x_{ut},$$

$$l_{itv}^5 = x_i'p_i'p_{iv}x_{it}, \quad l_{itv}^6 = s_3p_i'p_{iv}p_{iv}\varepsilon_i^2x_{it}, \quad l_{iwv}^7 = x_i'x_i'p_{iv}p_{iw},$$

55
Taking expectations, we can get expressions for $M^t$, $M^u$, $M^{tu}$, $M^{tv}$ and $M^{vw}$.

\[
M^t \equiv E[M^t_i] = \begin{pmatrix} 0 & 0 \\ 0 & 2l^1_t \end{pmatrix},
\]

\[
M^u \equiv E[M^u_i] = \begin{pmatrix} 0 & 2l^2_u \\ 2l^2_u & l^3_u \end{pmatrix},
\]

\[
M^{tu} \equiv E[M^{tu}_i] = \begin{pmatrix} 0 & 0 \\ 0 & -2l^4_{tu} \end{pmatrix},
\]

\[
M^{tv} \equiv E[M^{tv}_i] = \begin{pmatrix} 0 & -2l^5_{tv} \\ -2l^5_{tv} & -3l^6_{tv} \end{pmatrix},
\]

\[
M^{vw} \equiv E[M^{vw}_i] = \begin{pmatrix} -2l^7_{vw} & -3l^8_{vw} \\ -3l^8_{vw} & l^9_{vw} \end{pmatrix},
\]

where

\[
l^1_t = E[p_i'p_i\varepsilon_i x_{it}], \quad l^2_u = E[x_i'x_{it}p_{iw}\varepsilon_i], \quad l^3_u = E[p_i'p_i s_3 p_{iw}\varepsilon_i^3],
\]

\[
l^4_{tu} = E[p_i'x_{it} x_{ut}], \quad l^5_{tv} = E[x_i'p_{iv} p_{iw} x_{ut}], \quad l^6_{tv} = s_3 E[p_i'p_i p_{iw}\varepsilon_i^5 x_{ut}],
\]

\[
l^7_{vw} = E[x_i'x_{iv} p_{iw}], \quad l^8_{vw} = s_3 E[x_i'p_{iv} p_{iw}\varepsilon_i^3], \quad l^9_{vw} = s_4 E[p_i'p_i p_{iw}\varepsilon_i^4].
\]

With the assumption of linearity, higher order mean and variance in equations (5) and (6) can be written in terms of expectations that can be estimated in a monte carlo study. The calculations are relatively tedious and left to the Subsection 2.8.3.

### 2.6 Monte Carlo Experiments

In this section we analyze the performance of MSE criterion with monte carlo experiments. The data generating setup is similar to Donald and Newey [2001] and Donald
et al. [2002]:

\[ y_i = \beta_0 x_i + \varepsilon_i \]
\[ x_i = p'_i \pi + \eta_i \]

(8)

for \( i = 1, \ldots, n \). Therefore for \( K \) instruments, the following vector valued moment function is implied:

\[ g_i (\beta) = \begin{pmatrix} p_{1i} \\ p_{2i} \\ \vdots \\ p_{Ki} \end{pmatrix} (y_i - \beta x_i) \]

The main question we are interested is how to choose the number of instruments optimally. Without loss of generality we set \( \beta_0 = 0 \). The error terms are assumed to have the following variance-covariance structure:

\[ E \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \begin{pmatrix} \varepsilon_i & \eta_i \end{pmatrix} = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \]

(9)

Independent of the error terms, the instruments are drawn from a multivariate standard normal distribution: \( p_i \sim N (0, I_{K}) \) where \( K \) is the maximal number of instruments considered. As shown in Hahn and Hausman [2002], these assumptions imply that the first stage \( R^2 \) can be written as:

\[ R_j^2 = \frac{\pi' \pi}{\pi' \pi + 1} \]

As in Donald et al. [2002] we assume the true values of \( \pi_k \) coefficients in equation (9) follow a declining form:

\[ \pi_k = c (\bar{K}) \left( 1 - \frac{k}{\bar{K} + 1} \right)^4 \text{ for } k = 1, \ldots, \bar{K} \]
where constant \( c(\bar{K}) \) is chosen so that \( \pi'\pi = R^2_f/(1 - R^2_f) \). This represents a situation where one has prior information about the ranking of the relative relevance of the estimators. In this setting, if one is concerned about asymptotic efficiency, all of the potential \( \bar{K} \) instruments should be used. However, choosing the number of estimators that minimizes the higher order MSE will provide us a different instrument selection criteria. We report results for CUE which is a special case of GEL and which has interpretation in terms of GMM estimator as mentioned in Section 2. We report summary statistics for CUE estimator that uses all available instruments as well as CUE estimator when the number of instruments is chosen to minimize higher order MSE.

To make comparisons with Donald et al. [2002] straightforward, simulation parameters that are identical to their study are used. Two different sample sizes are considered: \( N = 200 \) or 1000. When the sample size is 200, \( R^2_f = 0.1 \) and \( \bar{K} = 10 \) are assumed and 500 replications are performed. In the larger sample size, \( R^2_f = 0.1 \) and \( \bar{K} = 20 \) are assumed and due to time constraints 200 replications are performed. The choice of \( R^2_f \) reflects the common weak instrument problem whereas the two selected \((N, \bar{K})\) pairs reflect the tendency of empirical researchers to use more moment conditions to improve efficiency as the number of observations grow. For each of these cases we consider

\[ c \in \{0.1, 0.5, 0.9\} \]

In addition, different from Donald et al. [2002], the impact of heteroskedasticity is considered. In this case, instead of equation (9), error terms have the following variance-covariance structure:

\[
E \left( \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \begin{pmatrix} \varepsilon_i & \eta_i \end{pmatrix} \right) = \begin{pmatrix} 1 + 0.3p^2_{i1} & c \\ c & 1 \end{pmatrix}
\]

where \( E(\varepsilon_i^2) = 1 + 0.3p^2_{i1} \) is chosen to provide a commonly observed amount of divergence between heteroskedasticity robust standard errors and standard errors
that are calculated with homoskedasticity assumption. Two different models, three different choices for residual correlations and homoskedastic and heteroskedastic cases result in 12 specifications.

The estimator that uses all instruments is indicated by $CUE_{all}$ whereas the estimator that uses a number of instruments that minimizes the MSE is indicated by $CUE_{op}$. The preliminary estimates of the objects that appear in estimated MSE were estimated by using the first and the strongest instrument.

As in Donald et al. [2002], we present robust measures of central tendency and dispersion. The median bias (Med. Bias), the median of the absolute deviations (MAD) of the estimator from the true value $\beta_0 = 0$ are computed. Dispersion is examined by looking at the difference between the 0.1 and 0.9 quantile (Dec. Rge) in the distribution of each estimator as well as standard deviation (std dev.). Standard deviation is not a robust measure of dispersion however it gives some indication of the presence of extreme outliers in finite samples. In addition, some summary statistics concerning the choice of number of instruments that minimizes the MSE are also reported. Compared to results in Donald et al. [2002] more dispersion in the optimal number of instruments is observed.

Tables 2.1-2.6 and 2.7-2.12 contains the summary statistics for the estimators in small samples ($N = 200$) and large samples ($N = 1000$). The results are in line with Donald et al. [2002]. The most significant impact in using optimal number of instruments instead of all instruments is a decline in the dispersion of the estimators as measured by the reported range and standard deviation measures. There is also a decline in the median of the absolute deviations. Similar to results in Donald et al. [2002] these improvements provided by choosing the number of instruments optimally are more pronounced when there is low to moderate degree of endogeneity. Finally, the improvements generated by choosing the number of instruments optimally is robust to heteroskedasticity.

Table 2.13 presents the summary statistics for the optimal number of instruments

---

5I thank Whitney Newey for pointing this out. For some examples on the impact of heteroskedasticity on the divergence between robust and non-robust standard errors see Wooldridge [2005].
across the replications for twelve different specifications possible. Similar to Donald et al. [2002], optimal number of instruments increases as endogeneity increases. However, optimal number of instruments is more dispersed compared to the results in Donald et al. [2002].

2.7 Conclusion

This paper proposes a criteria to choose the number of instruments in GEL estimation problems with heteroskedasticity. Similar to Donald et al. [2002], expressions to estimate the MSE that would result from choosing any number of instruments are derived and optimal number of instruments are chosen to minimize the MSE. The results of the simulation suggests that choosing the number of instruments based on higher order MSE reduces dispersion and bias of the estimator in a general scenario where there is heteroskedasticity and nonzero conditional third moments.

2.8 Appendix

2.8.1 Assumptions for the Stochastic Expansion of the GEL Estimator

**Assumption 1.** (a) $\beta_0 \in B$ is the unique solution to $E[g(z, \beta)] = 0$;
(b) $B$ is compact; (c) $g(z, \beta)$ is continuous at each $\beta \in B$ with probability one;
(d) $E[\sup_{\beta \in B} \| g(z, \beta) \|^\alpha] < \infty$ for some $\alpha > 2$; (e) $\Omega = E[g_i(\beta_0)g_i(\beta_0)']$ is nonsingular; (f) $s(v)$ is twice continuously differentiable in a neighborhood of zero.

**Assumption 2.** (a) $\beta_0 \in \text{int}(B)$; (b) $g(z, \beta)$ is continuously differentiable in a neighborhood $N$ of $\beta_0$ and $E[\sup_{\beta \in N} \| \partial g_i(\beta) / \partial \beta' \|] < \infty$;
(c) rank $(E[\partial g_i(\beta_0) / \partial \beta']) = p$.

**Assumption 3.** There is $b(z)$ with $E[b(z)^6] < \infty$ such that for $0 \leq j \leq 4$ and all $z$, $\nabla^j g(z, \beta)$ exists on a neighborhood $N$ of $\beta_0$, $\sup_{\beta \in N} \| \nabla^j g(z, \beta) \| \leq b(z)$, and for each
\( \beta \in N, \| \nabla^4 g(z, \beta) - \nabla^4 g(z, \beta_0) \| \leq b(z) \| \beta - \beta_0 \|, \) is four times continuously differentiable with Lipschitz fourth derivative in neighborhood of zero.

### 2.8.2 Appendix for Section 2.4

Since we assume that observations are independent:

\[
E \left[ \hat{M} \hat{U} \right] = E (M_i U_i) / n
\]

\[
\sum_{r=1}^{k} M^r E \left[ \hat{U}_r \hat{U}_r \right] / 2 = \sum_{r=1}^{k} M^r V_r / (2n)
\]

Where \( V_r \) denotes \( r \)th column of \( V \).

Calculating \( A \) and \( B \) are more involved and requires the following notation:

\[
\hat{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} Y_{ij} / n^2, \quad Y_{ij} = -M^{-1} (Y_{ij}^1 + Y_{ij}^2),
\]

\[
Y_{ij}^1 = (M_i - M) U_j, \quad Y_{ij}^2 = \sum_{r=1}^{k} U_i M^r U_j / 2
\]

\[
\hat{Z} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} Z_{ijk} / n^3, \quad Z_{ijk} = -M^{-1} \sum_{l=1}^{k} Z_{ijk}^l
\]

\[
Z_{ijk}^1 = (M_i - M) Y_{jk}, \quad Z_{ijk}^2 = \sum_{r=1}^{k} U_i M^r Y_{jk} / 2, \quad Z_{ijk}^3 = \sum_{r=1}^{k} Y_{ijr} M^r U_k / 2,
\]

\[
Z_{ijk}^4 = \sum_{r=1}^{k} U_{ir} (M^r_j - M^r) U_k / 2, \quad Z_{ijk}^5 = \sum_{r=1}^{k} \sum_{s=1}^{k} U_{ir} U_{js} M^{rs} U_k / 6
\]

where \( Y_{ijr} (U_{ir}) \) is the \( r \)th element of \( Y_{ij} (U_i) \).

**Calculating \( A \):** Assuming independent observations will lead to following conclusions: \( E [Y_{ij}] = 0 \) for \( i \neq j \) and \( E [Y_{ij} Y_{kl}] = 0 \) if any one of the indices is not equal to
any of the others. Therefore:

\[
\hat{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{Y_{ij}}{n^2} - \frac{E[Y_{ii}]/n}{n}
\]

Assuming that \(Y_{ii}\) has second moment, we have:

\[
E\left[ \hat{Y} \hat{Y}' \right] = \sum_{i,j,k,l=1}^{n} E\left[ Y_{ij} Y'_{kl} \right]/n^4 - E\left[ Y_{ii} \right] E\left[ Y'_{ii} \right]/n^2
\]

Using the above conclusions derived from independence:

\[
E\left[ \hat{Y} \hat{Y}' \right] = \sum_{i} E\left[ Y_{ii} Y'_{ii} \right]/n^4 + \sum_{i \neq j} (E\left[ Y_{ii} Y'_{jj} \right] + E\left[ Y_{ij} Y'_{ij} \right] + E\left[ Y_{ij} Y'_{ji} \right]) /n^4
\]

\[
- E\left[ Y_{ii} \right] E\left[ Y'_{ii} \right]/n^2
\]

\[
= E\left[ Y_{ii} Y'_{ii} \right]/n^3 + (E\left[ Y_{ii} Y'_{jj} \right] + E\left[ Y_{ij} Y'_{ij} \right] + E\left[ Y_{ij} Y'_{ji} \right]) (n^2 - n)/n^4
\]

\[
- E\left[ Y_{ii} \right] E\left[ Y'_{ii} \right]/n^2
\]

\[
= \left( E\left[ Y_{ij} Y'_{ij} \right] + E\left[ Y_{ij} Y'_{ji} \right] \right)/n^2 + O_p\left( n^{-3} \right)
\]

Hence,

\[
A = E\left[ Y_{ij} Y'_{ij} \right] + E\left[ Y_{ij} Y'_{ji} \right]
\]

\[
= M^{-1} \left( \Lambda_{b}^{11} + \Lambda_{b}^{22} + \Lambda_{b}^{12} + \Lambda_{c}^{12} + \Lambda_{c}^{11} + \Lambda_{c}^{22} + \Lambda_{c}^{12} + \Lambda_{c}^{22} \right) M^{-1'}
\]

where

\[
\Lambda_{b}^{11} = E\left[ Y_{ij} Y'_{ij} \right] = E\left[ (M_i - M) U_j V_j (M_i - M)' \right]
\]

\[
= E\left[ (M_i - M) V (M_i - M)' \right] = E\left[ M_i VM_i' \right] - VMVM',
\]

\[
\Lambda_{b}^{22} = E\left[ Y_{ij}^2 Y_{ij}^2 \right] = \sum_{r=1}^{k} \sum_{s=1}^{k} E\left[ U_{is} M_j V_j M_s M_j' U_{is} \right] /4
\]

\[
= \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs} M_j V M_s /4
\]

62
where $V_{rs}$ denotes the element at the $r^{th}$ row and $s^{th}$ column of $V$.

$$
\Lambda_0^{12} = E[Y_{ij}Y_{j'i'}] = \sum_{r=1}^{k} E[(M_i - M)U_jU_{j'}M'U_{ir}]/2
$$

$$
= \sum_{r=1}^{k} E[U_{ir}M_i]VM'/2,
$$

$$
\Lambda_c^{11} = E[Y_{ij}Y_{j'i'}] = E[(M_i - M)U_jU_{i'}(M_j - M)']
$$

$$
= E[M_iU_jU_{i'}M_{j'}],
$$

$$
\Lambda_c^{22} = E[Y_{ij}Y_{j'i'}] = \sum_{r=1}^{k} \sum_{s=1}^{k} E[U_{ir}M'rU_{i'}M'sU_{j'j}] /4
$$

$$
= \sum_{r=1}^{k} \sum_{s=1}^{k} M'rV_{jis}V_{is'M's'/4},
$$

$$
\Lambda_c^{12} = E[Y_{ij}Y_{j'i'}] = \sum_{r=1}^{k} E[(M_i - M)U_jU_{i'}M'rU_{j'r}] /2
$$

$$
= \sum_{r=1}^{k} E[M_iV_{is}U_{i'}M''] /2
$$

Calculating $B$:

Independent observations imply $E[U_iY_{jk}'] = 0$ if any of $i, j$ and $w$ is not equal to one of the others. Hence:

$$
E[\hat{Y}\hat{U}'] = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{w=1}^{n} E[Y_{ij}U_{k}'] /n^3 = E[Y_{ii'}U_{i}] /n^2
$$

$$
= -M^{-1}(D^1 + D^2)/n^2
$$
where

\[ D^1 = E \{ Y_{ii}' U_i' \} = E [(M_i - M) U_i U_i'] \]
\[ = E [M_i U_i U_i'] - M V, \]

\[ D^2 = E \{ Y_{ii}'^2 U_i' \} = \sum_{r=1}^{k} M^r E [U_{ir} U_i U_i'] / 2 \]

We also need \( E \{ \hat{Z} \hat{U}' \} : \)

\[ E \{ \hat{Z} \hat{U}' \} = \sum_{i,j,w,v=1}^{n} E \{ Z_{ijw} U_v' \} / n^4 = \sum_{i=1}^{n} E \{ Z_{ii} U_i' \} \]
\[ + \sum_{i\neq j} (E \{ Z_{ii} U_i' \} + E \{ Z_{ij} U_j' \} + E \{ Z_{ij} U_i' \}) / n^4 \]
\[ = (E \{ Z_{ii} U_i' \} + E \{ Z_{ij} U_j' \} + E \{ Z_{ij} U_i' \}) / n^2 + o (n^{-2}) \]
\[ = -M^{-1} \sum_{i=1}^{5} (F_a^i + F_b^i + F_c^i) / n^2 + o (n^{-2}) \]

where \( F_a^i = E \{ Z_{ii}^i U_j' \} , F_b^i = E \{ Z_{ij} U_j' \} , F_c^i = E \{ Z_{ij} U_i' \} \)

We need to write these terms in more open form in order to be able to implement a monte carlo study:

\[ F_a^1 = E \{ Z_{ii}^1 U_j' \} = -E \{ (M_i - M) M^{-1} (Y_{ij}^1 + Y_{ij}^2) U_j' \} \]
\[ = -E \{ (M_i - M) M^{-1} (M_i - M) U_j U_j' \} \]
\[ -E \left( (M_i - M) M^{-1} \sum_{r=1}^{k} U_{ir} M^r U_j U_j' / 2 \right) \]
\[ = - \left\{ E \{ M_i M^{-1} M_i \} - M + \sum_{r=1}^{k} E [U_{ir} M_i] M^{-1} M^r \right\} V, \]
\[ F_a^2 = E [Z_{ij}^2 U_j'] = - \sum_{r=1}^{k} E \left[ U_{ir} M^r M^{-1} (Y_{ri}^1 + Y_{ri}^2) U_j' \right] / 2 \]
\[ = - \sum_{r=1}^{k} E \left[ U_{ir} M^r M^{-1} \left\{ (M_i - M) U_j + \sum_{s=1}^{k} U_{is} M^s U_j / 2 \right\} U_j' \right] / 2 \]
\[ = - \left\{ \sum_{r=1}^{k} M^r M^{-1} E [U_{ir} M_i] / 2 + \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs} M^r M^{-1} M^s / 4 \right\} V, \]

\[ F_a^3 = E [Z_{ij}^3 U_j'] = \sum_{r=1}^{k} E \left[ Y_{iir} M^r U_j U_j' \right] / 2 = \sum_{r=1}^{k} E [Y_{iir} M^r V / 2, \]

\[ F_a^4 = E [Z_{ij}^4 U_j'] = \sum_{r=1}^{k} E \left[ U_{ir} (M_i^r - M^r) U_j U_j' \right] / 2 = \sum_{r=1}^{k} E [U_{ir} M_i^r] V / 2, \]

\[ F_a^5 = E [Z_{ij}^5 U_j'] = \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs} M^{rs} V / 6, \]

\[ F_b^1 = E [Z_{ijj}^1 U_j'] = - E \left[ (M_i - M) M^{-1} (Y_{ji}^1 + Y_{ji}^2) U_j' \right] \]
\[ = - E \left[ (M_i - M) M^{-1} (M_j - M) U_i U_j' \right] - E \left[ M_i M^{-1} \sum_{r=1}^{k} U_{jr} M^r U_i U_j' / 2 \right] \]
\[ = - E \left[ M_i M^{-1} M_j U_i U_j' \right] - \sum_{r=1}^{k} E \left[ M_i M^{-1} M^r U_i \right] V'_r / 2, \]

\[ F_b^2 = E [Z_{ijj}^2 U_j'] = - \sum_{r=1}^{k} E \left[ U_{ir} M^r M^{-1} (Y_{ji}^1 + Y_{ji}^2) U_j' \right] \]
\[ = - \sum_{r=1}^{k} E \left[ U_{ir} M^r M^{-1} \left\{ (M_j - M) U_i + \sum_{s=1}^{k} U_{js} M^s U_i / 2 \right\} U_j' \right] / 2 \]
\[ = - \sum_{r=1}^{k} M^r M^{-1} E [M_j V_{i} U_j'] / 2 - \sum_{r=1}^{k} \sum_{s=1}^{k} M^r M^{-1} M^s V_{i} V'_s / 4, \]

\[ F_b^3 = E [Z_{ijj}^3 U_j'] = \sum_{r=1}^{k} E \left[ Y_{iij} M^r U_i U_j' \right] / 2, \]

\[ F_b^4 = E [Z_{ijj}^4 U_j'] = \sum_{r=1}^{k} E \left[ U_{ir} (M_j^r - M^r) U_i U_j' \right] / 2 = \sum_{r=1}^{k} E [M_j^r V_{i} U_j'] / 2, \]
\[
F_b^6 = E \left[ Z_{ij}^5 U_i' \right] = \sum_{r=1}^{k} \sum_{s=1}^{k} E \left[ U_{ir} U_{js} M^{rs} U_i U_j' \right] / 6 = \sum_{r=1}^{k} \sum_{s=1}^{k} M^{rs} V_r V_s' / 6,
\]

\[
F_c^1 = E \left[ Z_{ij}^1 U_i' \right] = E \left[ (M_i - M) Y_{ij} U_i' \right] = E \left[ M_i E \left[ Y_{ij} \right] U_i' \right],
\]

\[
F_c^2 = E \left[ Z_{ij}^2 U_i' \right] = \sum_{r=1}^{k} E \left[ U_{ir} M^r Y_{ij} U_j' \right] / 2 = \sum_{r=1}^{k} M^r E \left[ Y_{ij} \right] V_r' / 2,
\]

\[
F_c^3 = \sum_{r=1}^{k} E \left[ Y_{ijr} M^r U_j U_i' \right] / 2,
\]

\[
F_c^4 = E \left[ Z_{ij}^4 U_i' \right] = \sum_{r=1}^{k} E \left[ U_{ir} (M_j^r - M^r) U_j U_i' \right] / 2 = \sum_{r=1}^{k} M_j^r U_j' V_r' / 2,
\]

\[
F_c^5 = E \left[ Z_{ij}^5 U_i' \right] = \sum_{r=1}^{k} \sum_{s=1}^{k} E \left[ U_{ir} U_{js} M^{rs} U_j U_i' \right] / 6 = \sum_{r=1}^{k} \sum_{s=1}^{k} M^{rs} V_r V_s' / 6
\]

### 2.8.3 Appendix for Section 2.5

Before deriving explicit expressions for higher order mean and variance of \( \hat{\theta}_{GEL} \), we need to do some preliminary calculations and introduce some additional notation. In particular, it would be helpful to write \( M^{-1}, U_i, \) and \( V \) in terms of partitioned matrices. Applying inverse of a partitioned matrix formula to equation (7) one would get:

\[
M^{-1} = \begin{pmatrix}
\Sigma & -H \\
-H' & -F
\end{pmatrix}
\]

where \( \Sigma = (K'\Omega^{-1}K)^{-1} \), \( H = -\Sigma K'\Omega^{-1} \) and \( F = \Omega^{-1} - \Omega^{-1} K \Sigma K'\Omega^{-1} \).

Therefore, \( U_i \) can be written as:

\[
U_i = -M^{-1} m_i (\theta_0) = \begin{pmatrix}
\Sigma & -H \\
-H' & -F
\end{pmatrix}
\begin{pmatrix}
0 \\
p_i \varepsilon_i
\end{pmatrix} = - \begin{pmatrix}
Hp_i \varepsilon_i \\
Fp_i \varepsilon_i
\end{pmatrix}
\]
Hence, \( V \) is in the following block diagonal form:

\[
V = E \left[ \begin{pmatrix} H \\ F \end{pmatrix} g_i (\beta_0) g'_i (\beta_0) \begin{pmatrix} H' \\ F' \end{pmatrix} \right] = \begin{pmatrix} H \\ F \end{pmatrix} \Omega \begin{pmatrix} H' \\ F' \end{pmatrix}
\]

\[
= \begin{pmatrix} H\Omega H' & H\Omega F' \\ F\Omega H' & F\Omega F' \end{pmatrix} = \begin{pmatrix} \Sigma & 0 \\ 0 & F \end{pmatrix}
\]

(11)

We also have:

\[
E [U_t U_i] = \begin{pmatrix} \Sigma, & 0 \\ 0, & F_v \end{pmatrix}, \quad E [U_w U_i] = \begin{pmatrix} 0 \\ F_v \end{pmatrix},
\]

\[
\hat{U} = - \begin{pmatrix} H \\ F \end{pmatrix} \sum_{i=1}^{n} p_i \epsilon_i
\]

where \( U_t (U_w) \) denotes the \( t^{th} \) \( (v^{th}) \) element of \(-H p_i \epsilon_i, -F p_i \epsilon_i\). Now, we can state implementable expressions for higher order mean and variance of \( \hat{\theta}_{GEL} \).

**Higher Order Bias of \( \hat{\theta}_{GEL} \)**

Equation (5) can we written as:

\[
BIAS \left( \hat{\theta}_{GEL} \right) = -M^{-1} \left[ E \left[ \hat{M} \hat{U} \right] + \sum_{i=1}^{p} M^t E \left[ \hat{U}_i \hat{U} \right] /2 + \sum_{v=1}^{m} M^v E \left[ \hat{U}_v \hat{U} \right] /2 \right]
\]

assuming independence of across observations this reduces to:

\[
BIAS \left( \hat{\theta}_{GEL} \right) = -M^{-1} \left[ B_1 + \sum_{v=1}^{m} \left( \begin{pmatrix} 0 & 2l^2_v \\ 0 & l^3_v \end{pmatrix} \right) \left( \begin{pmatrix} 0 \\ F_v \end{pmatrix} \right) /2 \right] /n
\]

\[
= -M^{-1} \left[ b_1 + b_2 \right] /n
\]

where

\[
b_1 = E \left[ M_t U_i \right], \quad b_2 = \sum_{v=1}^{m} M^v V_v /2
\]

and where \( V_v \) denotes \((p + v)^{th}\) column of \( V \).
Higher Order of Variance of $\hat{\theta}_{GEL}$

Equation (6) and the basic forms of $A$ and $B$ stated in Subsection 2.8.2 will be used to calculate the variance. Here, we will derive expressions that can easily be calculated in a monte carlo study and that will be used in calculating $A$ and $B$.

Calculating $A$: Equation (10) will be used to calculate $A$. Here are the expressions we need:

$$\Lambda_{b}^{11} = E[M_{i}VM_{i}'] - MM' = \delta_{1} - MM', $$

where

$$\delta_{1} = E[M_{i}VM_{i}'], $$

$$\Lambda_{b}^{22} = \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs}MM'/4 + \sum_{t=1}^{p} \sum_{u=1}^{p} V_{tu}M^{t}VM^{u}/4 + \sum_{v=1}^{m} \sum_{t=1}^{m} V_{vt}M^{v}VM^{t}/4 + \sum_{v=1}^{m} \sum_{u=1}^{m} V_{vu}M^{v}VM^{u}/4 $$

where $V_{tu}$ is the element at the $t^{th}$ row, $u^{th}$ column of $V$, $V_{tv}$ is the element at $t^{th}$ row $(p + v)^{th}$ column of $V$ and $V_{vu}$ is the element at the $(p + v)^{th}$ row and $(p + w)^{th}$ column of $V$. Since from equation (11) $V$ is block diagonal:

$$\Lambda_{b}^{22} = \delta_{2} + \delta_{3}, $$

where

$$\delta_{2} = \sum_{t=1}^{p} \sum_{u=1}^{p} \sum_{v=1}^{m} \sum_{w=1}^{m} \Sigma_{tu}M^{t}VM^{w}/4, \quad \delta_{3} = \sum_{v=1}^{m} \sum_{w=1}^{m} F_{vw}M^{v}VM^{w}/4 $$

$$\Lambda_{b}^{12} = \sum_{r=1}^{k} E[U_{r}M_{i}] VMM'/2 = \delta_{4} + \delta_{5}, $$
where
\[\delta_4 = \sum_{t=1}^{p} E[U_{it}M_t] VM^t/2, \quad \delta_5 = \sum_{ve=1}^{m} E[U_{iv}M_i] VM^v/2 \]
\[\Lambda_{c}^{11} = E[M_iU_jU_i^tM_j']\]

after some algebra, this can be written as\textsuperscript{6}:
\[\Lambda_{c}^{11} = \begin{pmatrix} \delta_6 & \delta_9 - \delta_{10} \\ \delta_7 - \delta_8 & \delta_{11} - \delta_{12} - \delta_{13} + \delta_{14} \end{pmatrix},\]

where
\[\delta_6 = E[p_jF_j\xi F_{j'k}x_j\varepsilon_j], \quad \delta_7 = E[p_{11}H_j\xi F_{j'k}x_{k'}\varepsilon_{k'}], \quad \delta_8 = E[p_jp_j'F_j\xi F_{j'k}x_{k'}\varepsilon_{k'}^3],\]
\[\delta_9 = E[p_jF_j\xi H'_{j'k}x_j\varepsilon_j], \quad \delta_{10} = E[p_jF_j\xi F_{j'k}p_{j'k'}\varepsilon_{j'k'}^3], \quad \delta_{11} = \xi E[H'_{j'k}(H_{j'k})x_j\varepsilon_j],\]
\[\delta_{12} = \xi E[F_{j'k}p_{j'k'}\varepsilon_{j'k'}^3(H_{j'k})], \quad \delta_{13} = E[p_{11}p_jF_j(H_{j'k})\varepsilon_{j'k'}^3], \quad \delta_{14} = E[p_jp_{11}F_{j'k}p_{j'k'}F_{j'k}p_{j'k'}p_{j'k'}\varepsilon_{j'k'}^3\varepsilon_{k'}^3],\]

and
\[\xi = E[p_jp_j'x_j\varepsilon_j]\]
\[\Lambda_{c}^{22} = \sum_{r=1}^{k} \sum_{s=1}^{k} M^r V_{rs} V_{st} M^{st}/4 = 2\delta_{15} + \delta_{16},\]

where
\[\delta_{15} = \sum_{t=1}^{p} \sum_{ve=1}^{m} M^t V_{ve} V_{ve} M^{ve}/4, \quad \delta_{16} = \sum_{ve=1}^{m} \sum_{w=1}^{m} M^v V_{we} V_{we} M^{vw}/4\]
\[\Lambda_{c}^{12} = \sum_{r=1}^{k} E[M_iV_{r}U_i' M^t]/2 = \delta_{17} + \delta_{18},\]

where
\[\delta_{17} = \sum_{t=1}^{p} E[M_iV_{r}U_i' M^t]/2, \quad \delta_{18} = \sum_{ve=1}^{m} E[M_iV_{ve}U_i' M^{ve}]/2\]

\textsuperscript{6}In the monte carlo study, we assume \(p = 1\), this helps in calculating expectations that involves cross products from different time periods. Here, for \(\Lambda_{c}^{11}\) and \(F_{c}^{1}\) the provided expressions are simplified by assuming \(p = 1\). More general expressions are available upon request.
Calculating $B$ : Formulas in Subsection 2.8.2 implies:

$$B = -M^{-1} (D^1 + D^2) - M^{-1} \sum_{l=1}^{5} (F^l_a + F^l_b + F^l_c)$$

Here, we will provide explicit expressions to calculate $B$:

$$D^1 = E[M_iU_i^tU_i'] - MV = \delta_{19} - MV,$$

where

$$\delta_{19} = E[M_iU_i^tU_i']$$

$$D^2 = \sum_{r=1}^{k} M^r E[U_{ir}U_i^tU_i'] / 2 = \delta_{20} + \delta_{21},$$

where

$$\delta_{20} = \sum_{t=1}^{p} M^t E[U_{it}U_i^tU_i'] / 2, \quad \delta_{21} = \sum_{v=1}^{m} M^v E[U_{iv}U_i^tU_i'] / 2$$

$$F^1_a = \left\{ E[M_iM^{-1}M_i] - M + \sum_{r=1}^{k} E[U_{ir}M_i] M^{-1}M^r/2 \right\} V$$

$$= \{ \delta_{22} - M + \delta_{23} + \delta_{24} \} V,$$

where

$$\delta_{22} = E[M_iM^{-1}M_i]$$

$$\delta_{23} = \sum_{t=1}^{p} E[U_{it}M_i] M^{-1}M^t/2, \quad \delta_{24} = \sum_{v=1}^{m} E[U_{iv}M_i] M^{-1}M^v/2$$

$$F^2_a = -\left\{ \sum_{r=1}^{k} M^r M^{-1}E[U_{ir}M_i] / 2 + \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs}M^r M^{-1}M^s / 4 \right\} V$$

noting that $V$ is block diagonal, this can be written as:

$$= \{ \delta_{23} + \delta_{24} + \delta_{25} + \delta_{26} \} V,$$
where
\[
\delta_{25} = \sum_{t=1}^{p} \sum_{u=1}^{p} \Sigma_{tu} M^t M^{-1} M^u / 4, \quad \delta_{26} = \sum_{v=1}^{m} \sum_{w=1}^{m} V_{vw} M^v M^{-1} M^w / 4
\]
\[
F^3_a = \sum_{r=1}^{k} E [Y_{itr}] M^r V / 2 = \delta_{27} + \delta_{28},
\]
where
\[
\delta_{27} = \sum_{t=1}^{p} E [Y_{itt}] M^t V / 2, \quad \delta_{28} = \sum_{v=1}^{m} E [Y_{ivv}] M^v V / 2
\]
\[
F^4_a = \sum_{r=1}^{k} E [U_{itr} M^r] V / 2 = \{\delta_{29} + \delta_{30}\} V / 2,
\]
where
\[
\delta_{29} = \sum_{t=1}^{p} E [U_{itt} M^t], \quad \delta_{30} = \sum_{v=1}^{m} E [U_{ivv} M^v]
\]
\[
F^5_a = \sum_{r=1}^{k} \sum_{s=1}^{k} V_{rs} M^{rs} V / 6 = \delta_{31} + \delta_{32},
\]
where
\[
\delta_{31} = \sum_{t=1}^{p} \sum_{u=1}^{p} \Sigma_{tu} M^t u / 6, \quad \delta_{32} = \sum_{v=1}^{m} \sum_{w=1}^{m} F_{vw} M^v w V / 6
\]
\[
F^1_b = -E [M_i M^{-1} M_j U_i U_j'] - \sum_{r=1}^{k} E [M_i M^{-1} M^r U_i] V'_r / 2
\]
\[
= -\delta_{33} - \delta_{34} - \delta_{35},
\]
where
\[
\delta_{33} = E [M_i M^{-1} M_j U_i U_j']
\]
\[
\delta_{34} = \sum_{t=1}^{p} E [M_i M^{-1} M^t U_i] V'_t / 2, \quad \delta_{35} = \sum_{v=1}^{m} E [M_i M^{-1} M^v U_i] V'_v / 2
\]

Since \(\delta_{33}\) involves expectations of cross-products, we need to rewrite it in a form that
would make the calculations easier:

\[
\delta_{33} = E \left[ \begin{pmatrix} 0 & -x_i p'_i \\ -p_i x'_i & p_i p'_i \varepsilon_i^2 \end{pmatrix} \begin{pmatrix} \Sigma & -H \\ -H' & -F \end{pmatrix} \begin{pmatrix} 0 & -x_j p'_j \\ -p_j x'_j & p_j p'_j \varepsilon_j^2 \end{pmatrix} \begin{pmatrix} H \\ F \end{pmatrix} p_i \varepsilon_i \varepsilon_j p'_j \left( H' \quad F' \right) \right]
\]

after a number of steps \( \delta_{33} \) can be written as:

\[
\begin{pmatrix}
(-\delta_9 - \delta_{36} + \delta'_6) H' \\
(-\delta_3 + \delta_{12} + \delta_{38} + \delta_{13} + \delta_{12}) H' - \delta_{39} \\
(-\delta_{37} + \delta'_{12} + \delta_{38} + \delta_{13} + \delta_{12} - \delta_{14}) F
\end{pmatrix}
\]

where

\[
\delta_{36} = H \xi F \xi, \quad \delta_{37} = \xi H' H \xi
\]

\[
\delta_{38} = \xi \Sigma F \xi, \quad \delta_{39} = E \left[ p_i p'_i F E \zeta F p_i \varepsilon_i^2 \right]
\]

and

\[
\zeta = E \left( p_j p'_j (H p_j) \varepsilon_j^3 \right)
\]

\[
F_b^2 = -\sum_{r=1}^{k} M^r M^{-1} E \left[ M_j V_r U'_j \right] /2 - \sum_{r=1}^{k} \sum_{s=1}^{k} M^r M^{-1} M^s V_r V'_s /4
\]

\[
= -\delta_{40} - \delta_{41} - \delta_{42} - \delta_{43} - \delta_{44} - \delta_{45}
\]

where

\[
\delta_{40} = \sum_{t=1}^{p} M^t M^{-1} E \left[ M_j V_t U'_j \right] /2, \quad \delta_{41} = \sum_{v=1}^{m} M^v M^{-1} E \left[ M_j V_v U'_j \right] /2
\]

\[
\delta_{42} = \sum_{t=1}^{p} \sum_{u=1}^{p} M^t M^{-1} M^u V_t V'_u /4, \quad \delta_{43} = \sum_{t=1}^{p} \sum_{v=1}^{m} M^t M^{-1} M^v V_t V'_v /4
\]

\[
\delta_{44} = \sum_{v=1}^{m} \sum_{t=1}^{p} M^v M^{-1} M^t V_v V'_t /4, \quad \delta_{45} = \sum_{v=1}^{m} \sum_{u=1}^{m} M^v M^{-1} M^u V_v V'_u /4
\]

72
\[ F^3_b = \sum_{r=1}^{k} E \left[ Y_{ijr} M^t U_i U'_j \right] / 2 \]
\[ = \sum_{t=1}^{p} E \left[ Y_{ijt} M^t U_i U'_j \right] / 2 + \sum_{v=1}^{m} E \left[ Y_{ijv} M^v U_i U'_j \right] / 2 \]

After a number of steps:

\[ = \delta_{46} + \delta_{47} + \delta_{48} + \delta_{50} + \delta_{51}, \]

where

\[
\delta_{46} = \sum_{t=1}^{p} M^t \left( H \delta_{t2}^t \delta_{53} H' \ H \delta_{52}^t \delta_{53} F \right) / 2,
\]
\[
\delta_{47} = \sum_{v=1}^{m} M^v \left( H \delta_{54}^v \delta_{53} H' \ H \delta_{54}^v \delta_{53} F \right) / 2,
\]
\[
\delta_{48} = \sum_{t=1}^{p} \sum_{u=1}^{p} M^t \left( H \delta_{55}^u \delta_{53} H' \ H \delta_{55}^u \delta_{53} F \right) / 4,
\]
\[
\delta_{49} = \sum_{t=1}^{p} \sum_{v=1}^{m} M^t \left( H \delta_{56}^u \delta_{53} H' \ H \delta_{56}^u \delta_{53} F \right) / 4,
\]
\[
\delta_{50} = \sum_{v=1}^{m} \sum_{t=1}^{p} M^v \left( H \delta_{57}^v M_i \delta_{53} H' \ H \delta_{57}^v M_i \delta_{53} F \right) / 4,
\]
\[
\delta_{51} = \sum_{v=1}^{m} \sum_{w=1}^{m} M^v \left( H \delta_{58}^w M^w \delta_{53} H' \ H \delta_{58}^w M^w \delta_{53} F \right) / 4,
\]

and

\[
\delta_{52}^t = -E \left( p_i \varepsilon_i \left( M^{-1} \right)_t \ M_i \right), \quad \delta_{53} = E \left( U_j p'_j \varepsilon_j \right),
\]
\[
\delta_{54}^v = -E \left( p_i \varepsilon_i \left( M^{-1} \right)_v \ M_i \right), \quad \delta_{55}^v = -E \left( p_i \left( M^{-1} \right)_t \ U_{iuv} \varepsilon_i \right),
\]
\[
\delta_{56}^v = -E \left( p_i \left( M^{-1} \right)_t \ U_{iuv} \varepsilon_i \right), \quad \delta_{57}^v = -E \left( p_i \left( M^{-1} \right)_v \ U_{itv} \varepsilon_i \right),
\]
\[
\delta_{58}^w = -E \left( p_i \left( M^{-1} \right)_w \ U_{iuv} \varepsilon_i \right)
\]
Here, \((M^{-1})_t\) denotes the \(t^{th}\) row \(M^{-1}\), whereas \((M^{-1})_v\) denotes the \((p + v)^{th}\) row of \(M^{-1}\). Similarly, \(U_{iu}\) denotes the \(u^{th}\) element of \(U_i\), whereas \(U_{iw}\) denotes the \((p + w)^{th}\) element of \(U_i\).

\[
F_b^4 = \sum_{r=1}^{k} E \left[ M_j^r V_r U'_j \right] / 2 = \delta_{59} + \delta_{60},
\]

where

\[
\delta_{59} = \sum_{t=1}^{p} E \left[ M_j^t V_t U'_j \right] / 2, \quad \delta_{60} = \sum_{v=1}^{m} E \left[ M_j^v V_v U'_j \right] / 2
\]

\[
F_b^5 = \sum_{r=1}^{k} \sum_{s=1}^{k} M^{rs} V_r V'_s / 6 = \delta_{61} + \delta_{62} + \delta_{63},
\]

where

\[
\delta_{61} = \sum_{t=1}^{p} \sum_{u=1}^{m} M^{tu} V_t V'_u / 6, \quad \delta_{62} = \sum_{v=1}^{m} \sum_{t=1}^{p} M^{tv} V_v V'_t / 6,
\]

\[
\delta_{63} = \sum_{u=1}^{m} \sum_{w=1}^{m} M^{uw} V_u V'_w / 6
\]

\[
F_c^2 = \sum_{r=1}^{k} M^r E [Y_{jj}] V'_r / 2 = \delta_{64} + \delta_{65},
\]

where

\[
\delta_{64} = \sum_{t=1}^{p} M^t E [Y_{jj}] V'_t / 2, \quad \delta_{65} = \sum_{v=1}^{m} M^v E [Y_{jj}] V'_v / 2
\]

\[
F_c^3 = \sum_{r=1}^{k} E \left[ Y_{ijr} M^r U_j U'_i \right] / 2
\]

After some algebra, it can be shown that \(F_c^3\) can be written as:

\[
= \delta_{66} + \delta_{67} + \delta_{68} + \delta_{69} + \delta_{70} + \delta_{71},
\]

74
where

\[ \delta_{66} = \sum_{t=1}^{p} M^t \left( \begin{array}{cc} H\delta_{53}^t \delta_{52}^t H' & H\delta_{53}^t \delta_{52}^t F \\ F\delta_{53}^t \delta_{52}^t H' & F\delta_{53}^t \delta_{52}^t F \end{array} \right) /2, \]

\[ \delta_{67} = \sum_{v=1}^{m} M^v \left( \begin{array}{cc} H\delta_{53}^v \delta_{54}^v H' & H\delta_{53}^v \delta_{54}^v F \\ F\delta_{53}^v \delta_{54}^v H' & F\delta_{53}^v \delta_{54}^v F \end{array} \right) /2, \]

\[ \delta_{68} = \sum_{t=1}^{p} \sum_{u=1}^{p} M^t \left( \begin{array}{cc} H\delta_{53}^t M^u \delta_{55}^u H' & H\delta_{53}^t M^u \delta_{55}^u F \\ F\delta_{53}^t M^u \delta_{55}^u H' & F\delta_{53}^t M^u \delta_{55}^u F \end{array} \right) /4, \]

\[ \delta_{69} = \sum_{t=1}^{m} \sum_{v=1}^{m} M^t \left( \begin{array}{cc} H\delta_{53}^t M^v \delta_{56}^v H' & H\delta_{53}^t M^v \delta_{56}^v F \\ F\delta_{53}^t M^v \delta_{56}^v H' & F\delta_{53}^t M^v \delta_{56}^v F \end{array} \right) /4, \]

\[ \delta_{70} = \sum_{v=1}^{m} \sum_{t=1}^{m} M^v \left( \begin{array}{cc} H\delta_{53}^v M^t \delta_{57}^t H' & H\delta_{53}^v M^t \delta_{57}^t F \\ F\delta_{53}^v M^t \delta_{57}^t H' & F\delta_{53}^v M^t \delta_{57}^t F \end{array} \right) /4, \]

\[ \delta_{71} = \sum_{v=1}^{m} \sum_{w=1}^{m} M^v \left( \begin{array}{cc} H\delta_{53}^v M^w \delta_{58}^w H' & H\delta_{53}^v M^w \delta_{58}^w F \\ F\delta_{53}^v M^w \delta_{58}^w H' & F\delta_{53}^v M^w \delta_{58}^w F \end{array} \right) /4, \]

\[ F_c^4 = \sum_{r=1}^{k} E \left[ M^r U_j \right] V'_r /2 = \delta_{72} + \delta_{73}, \]

where

\[ \delta_{72} = \sum_{t=1}^{p} E \left[ M^t U_j \right] V'_t /2, \quad \delta_{73} = \sum_{v=1}^{m} E \left[ M^v U_j \right] V'_v /2 \]

\[ F_c^5 = \sum_{r=1}^{k} \sum_{s=1}^{k} M^{rs} V_s V'_r /6 = \delta_{74} + \delta_{75} + \delta_{76}, \]

where

\[ \delta_{74} = \sum_{t=1}^{p} \sum_{v=1}^{m} M^{tv} V_t V'_v /6, \quad \delta_{75} = \sum_{v=1}^{m} \sum_{t=1}^{p} M^{tv} V_t V'_v /6, \quad \delta_{76} = \sum_{v=1}^{m} \sum_{w=1}^{m} M^{tv} V_w V'_v /6 \]
### 2.8.4 Simulation Results

Table 2.1: CUE Estimators: Small Sample, Weak Endogeneity and Homoskedasticity

\[ N = 200, \bar{K} = 10, Cov = 0.1 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CUEall</td>
<td>0.003</td>
<td>0.201</td>
<td>0.7981</td>
<td>0.8189</td>
</tr>
<tr>
<td>CUEop</td>
<td>0.005</td>
<td>0.175</td>
<td>0.658</td>
<td>0.5438</td>
</tr>
</tbody>
</table>

Table 2.2: CUE Estimators: Small Sample, Weak Endogeneity and Heteroskedasticity

\[ N = 200, \bar{K} = 10, Cov = 0.1 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CUEall</td>
<td>-0.007</td>
<td>0.233</td>
<td>0.9121</td>
<td>0.7644</td>
</tr>
<tr>
<td>CUEop</td>
<td>0.013</td>
<td>0.211</td>
<td>0.8201</td>
<td>0.6922</td>
</tr>
</tbody>
</table>
Table 2.3: CUE Estimators: Small Sample, Medium Endogeneity and Homoskedasticity

\[ N = 200, \bar{K} = 10, Cov = 0.5 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{\text{all}} )</td>
<td>0.001</td>
<td>0.169</td>
<td>0.678</td>
<td>0.3579</td>
</tr>
<tr>
<td>( CUE_{\text{op}} )</td>
<td>0.035</td>
<td>0.171</td>
<td>0.638</td>
<td>0.3458</td>
</tr>
</tbody>
</table>

Table 2.4: CUE Estimators: Small Sample, Medium Endogeneity and Heteroskedasticity

\[ N = 200, \bar{K} = 10, Cov = 0.5 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{\text{all}} )</td>
<td>0.007</td>
<td>0.183</td>
<td>0.7961</td>
<td>0.7738</td>
</tr>
<tr>
<td>( CUE_{\text{op}} )</td>
<td>0.043</td>
<td>0.179</td>
<td>0.744</td>
<td>0.7549</td>
</tr>
</tbody>
</table>

Table 2.5: CUE Estimators: Small Sample, Strong Endogeneity and Homoskedasticity

\[ N = 200, \bar{K} = 10, Cov = 0.9 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{\text{all}} )</td>
<td>-0.005</td>
<td>0.159</td>
<td>0.706</td>
<td>0.4799</td>
</tr>
<tr>
<td>( CUE_{\text{op}} )</td>
<td>0.015</td>
<td>0.155</td>
<td>0.68</td>
<td>0.3809</td>
</tr>
</tbody>
</table>
Table 2.6: CUE Estimators: Small Sample, Strong Endogeneity and Heteroskedasticity

\[ N = 200, K = 10, Cov = 0.9 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{all} )</td>
<td>-0.001</td>
<td>0.189</td>
<td>0.716</td>
<td>0.6268</td>
</tr>
<tr>
<td>( CUE_{op} )</td>
<td>0.027</td>
<td>0.181</td>
<td>0.688</td>
<td>0.6117</td>
</tr>
</tbody>
</table>

Table 2.7: CUE Estimators: Large Sample, Weak Endogeneity and Homoskedasticity

\[ N = 1000, K = 20, Cov = 0.1 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{all} )</td>
<td>-0.007</td>
<td>0.073</td>
<td>0.3</td>
<td>0.1145</td>
</tr>
<tr>
<td>( CUE_{op} )</td>
<td>-0.003</td>
<td>0.067</td>
<td>0.286</td>
<td>0.1095</td>
</tr>
</tbody>
</table>

Table 2.8: CUE Estimators: Large Sample, Weak Endogeneity and Heteroskedasticity

\[ N = 1000, K = 20, Cov = 0.1 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{all} )</td>
<td>-0.021</td>
<td>0.083</td>
<td>0.334</td>
<td>0.1252</td>
</tr>
<tr>
<td>( CUE_{op} )</td>
<td>-0.011</td>
<td>0.081</td>
<td>0.314</td>
<td>0.118</td>
</tr>
</tbody>
</table>
Table 2.9: CUE Estimators: Large Sample, Medium Endogeneity and Homoskedasticity

\[ N = 1000, \bar{K} = 20, \text{Cov} = 0.5 \]

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>0.005</td>
<td>0.063</td>
<td>0.252</td>
<td>0.1017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>0.013</td>
<td>0.065</td>
<td>0.234</td>
<td>0.0982</td>
</tr>
</tbody>
</table>

Table 2.10: CUE Estimators: Large Sample, Medium Endogeneity and Heteroskedasticity

\[ N = 1000, \bar{K} = 20, \text{Cov} = 0.1 \]

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>-0.007</td>
<td>0.089</td>
<td>0.312</td>
<td>0.1254</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>0.003</td>
<td>0.089</td>
<td>0.29</td>
<td>0.1232</td>
</tr>
</tbody>
</table>

Table 2.11: CUE Estimators: Large Sample, Strong Endogeneity and Homoskedasticity

\[ N = 1000, \bar{K} = 20, \text{Cov} = 0.9 \]

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>-0.003</td>
<td>0.071</td>
<td>0.272</td>
<td>0.107</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator ( \text{CUE}_{\text{all}} )</th>
<th>Med. Bias</th>
<th>Med. AD</th>
<th>Dec. Rge</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{CUE}_{\text{op}} )</td>
<td>0.005</td>
<td>0.067</td>
<td>0.264</td>
<td>0.1046</td>
</tr>
</tbody>
</table>
Table 2.12: CUE Estimators: Large Sample, Strong Endogeneity and Heteroskedasticity

\[ N = 1000, \bar{K} = 20, Cov = 0.9 \]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( CUE_{all} )</td>
<td>-0.013</td>
<td>0.077</td>
<td>0.304</td>
<td>0.1202</td>
</tr>
<tr>
<td>( CUE_{op} )</td>
<td>-0.005</td>
<td>0.077</td>
<td>0.294</td>
<td>0.1175</td>
</tr>
</tbody>
</table>
Table 2.13: Instrument Selection Based on MSE

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>Mode</th>
<th>1Q</th>
<th>Med.</th>
<th>3Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.1$, homoskedasticity</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.1$, heteroskedasticity</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.5$, homoskedasticity</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.5$, heteroskedasticity</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.9$, homoskedasticity</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$N = 200, \bar{K} = 10, Cov = 0.9$, heteroskedasticity</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.1$, homoskedasticity</td>
<td>20</td>
<td>10</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.1$, heteroskedasticity</td>
<td>20</td>
<td>10</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.5$, homoskedasticity</td>
<td>20</td>
<td>11</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.5$, heteroskedasticity</td>
<td>20</td>
<td>10</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.9$, homoskedasticity</td>
<td>20</td>
<td>15</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>$N = 1000, \bar{K} = 20, Cov = 0.9$, heteroskedasticity</td>
<td>20</td>
<td>13</td>
<td>17</td>
<td>19</td>
</tr>
</tbody>
</table>
Bibliography


Chapter 3

Inference on Sets in Finance

In this paper we introduce various set inference problems as they appear in finance and propose practical and powerful inferential tools. Our tools will be applicable to any problem where the set of interest solves a system of smooth estimable inequalities, though we will particularly focus on the following two problems: the admissible mean-variance sets of stochastic discount factors and the admissible mean-variance sets of asset portfolios. We propose to make inference on such sets using weighted likelihood-ratio and Wald type statistics, building upon and substantially enriching the available methods for inference on sets.

3.1 Introduction

Let us now introduce the first problem. We begin by recalling two equations used by Cochrane [2005] to effectively summarize the science of asset pricing:

\[
P_t = E_t[M_{t+1}X_{t+1}]
\]
\[
M_{t+1} = f(Z_{t+1}, \text{parameters}),
\]

where \(P_t\) is an asset price, \(X_{t+1}\) is the asset payoff, \(M_{t+1}\) is the stochastic discount factor (SDF) or pricing kernel (PK), which is a function \(f\) of some data \(Z_{t+1}\) and

\(^1\)This Chapter is the product of joint work with Victor Chernozhukov and Konrad Menzel.
parameters, and $E_t$ is the conditional expectation given information at time $t$. The set of SDFs $M_t$ that can price existing assets generally form a proper set, that is, a set that is not a singleton. SDFs are not unique, because the existing payoffs to assets do not span the entire universe of possible random payoffs. Dynamic asset pricing models provide families of potential SDFs, for example, the standard consumption model predicts that an appropriate SDF can be stated in terms of inter temporal marginal rate of substitution:

$$M_t = \beta \frac{u'(C_{t+1})}{u'(C_t)} ,$$

where $u$ denotes a utility function parameterized by some parameters, $C_t$ denotes consumption at time $t$, and $\beta$ denotes the subjective discount factor.

The basic econometric problem is to check which families of SDFs price the assets correctly and which do not. In other words, we want to check whether given families or sub families of SDFs are valid or not. One leading approach for performing the check is to see whether mean and standard deviation of SDFs

$$\{\mu_M, \sigma_M\}$$

are admissible. The set of admissible means and standard deviations

$$\Theta_0 := \{ \text{admissible pairs } (\mu, \sigma^2) \in \mathbb{R}^2 \cap K \},$$

which is introduced by Hansen and Jagannathan [1991] is known as the Hansen-Jagannathan set and the boundary of the set $\Theta_0$ is known as the Hansen-Jagannathan bound. In order to give a very specific, canonical example, let $v$ and $\Sigma$ denote the vector of mean returns and covariance matrix to assets 1, ..., $N$ which are assumed not to vary with information sets at each period $t$. Let us denote

$$A = v'\Sigma^{-1}v, B = v'\Sigma^{-1}1_N, C = 1'_n\Sigma^{-1}1_N$$

(3.1)
where $1_N$ is a column vector of ones. Then the minimum variance $\sigma^2(\mu)$ achievable by a SDF given mean $\mu$ of the SDF is equal to

$$\sigma^2(\mu) = (1 - \mu v)' \Sigma^{-1} (1 - \mu v) = A \mu^2 - 2B \mu + C$$

Therefore, the HJ set is equal to

$$\Theta_0 = \{ (\mu, \sigma) \in \mathbb{R}^2 \cap K : \sigma(\mu) - \sigma \leq 0 \},$$

where $K$ is any compact set. That is,

$$\Theta_0 = \{ \theta \in \Theta : m(\theta) \leq 0 \}.$$

Note that the inequality-generating function $m(\theta)$ depends on the unknown parameters, the means and covariance of returns, $m(\theta) = m(\theta, \gamma)$ and $\gamma = \text{vec}(v, \Sigma)$.

Let us now describe the second problem. The classical Markowitz [1952] problem is to minimize the risk of a portfolio given some attainable level of return:

$$\min_w E_t[\sigma^2_{\mu}(r_{p,t+1} - E_t[r_{p,t+1}])]^2 \text{ such that } E_t[r_{p,t+1}] = \mu,$$

where $r_{p,t+1}$ is portfolio return, determined as $r_{p,t+1} = w r_{t+1}$, where $w$ is a vector of portfolio “weights” and $r_{t+1}$ is a vector of returns on available assets. In a canonical version of the problem, we have that the vector of mean returns $v$ and covariance of returns $\Sigma$ do not vary with time period $t$, so that the problem becomes:

$$\sigma(\mu) = \min_w w' \Sigma w \text{ such that } w' v = \mu.$$

An explicit solution for $\sigma(\mu)$ takes the form,

$$\sigma^2(\mu) = \frac{C \mu^2 - 2B \mu + A}{AC - B^2}$$
where A, B and C are as in equation 3.1.

Therefore, the Markowitz (M) set of admissible standard deviations and means is given by

\[
\Theta_0 = \{(\mu, \sigma) \in \mathbb{R}^2 \cap K : \sigma(\mu, \sigma) < 0\},
\]

that is,

\[
\Theta_0 = \{\theta \in \Theta : m(\theta) \leq 0\}.
\]

The boundary of the set \(\Theta_0\) is known as the efficient frontier. Note that as in HJ example, the inequality-generating function \(m(\theta)\) depends on the unknown parameters, the means and covariance of returns, \(m(\theta) = m(\theta, \gamma)\), where \(\gamma = \text{vec}(v, \Sigma)\).

The basic problem of this paper is to develop inference methods on HJ and M sets, accounting for uncertainty in the estimation of parameters of the inequality-generating functions. The problem is to construct a confidence region \(R\) such that

\[
\lim_{n \to \infty} P\{\Theta_0 \subseteq R\} = 1 - \alpha.
\]

We will construct confidence regions for HJ sets using LR and Wald-type Statistics, building on and simultaneously enriching the approaches suggested in Chernozhukov et al. [2007], Beresteanu and Molinari [2008], and Molchanov [1998]. We also would like to ensure that confidence regions \(R\) are as small as possible and converge to \(\Theta_0\) at the most rapid attainable speed. We need the confidence region \(R\) for entire set \(\Theta_0\) in order to test validity of sets of SDFs. Once \(R\) is constructed, we can test infinite number of composite hypotheses, current and future, without compromising the significance level. Indeed, a typical application of HJ sets determines which sets of \((\mu, \sigma)'s\) within a given family fall in the HJ set and which do not. Similar comments about applicability of our approach go through for the M sets as well.

Our approach to inference using weighted Wald-type statistics complements and enriches the approach based on the directed Hausdorff distance suggested in Beresteanu and Molinari [2008] and Molchanov [1998]. By using weighting in the construction of the Wald-type statistics, we endow this approach with better invariance properties.
to parameter transformations, which results in noticeably sharper confidence sets, at least in the canonical empirical example that we will show. Thus, our construction is of independent interest for this type of inference, and is a useful complement to the work of Beresteanu and Molinari [2008] and Molchanov [1998]. Furthermore, our results on formal validity of the bootstrap for LR-type and W-type statistics are also of independent interest.

The rest of the paper is organized as follows. In Section 2 we present our estimation and inference results. In Section 3 we present an empirical example, illustrating the constructions of confidence sets for HJ sets. In Section 4 we draw conclusions and provide direction for further research. In the Appendix, we collect the proofs of the main results.

3.2 Estimation and Inference Results

3.2.1 Basic Constructions

The proofs of the provided results in this section can be found in Subsection 3.5.2. We first introduce our basic framework. We have an inequality-generating function:

\[ m: \Theta \mapsto \mathbb{R}. \]

The set of interest is the solution of the inequalities generated by the function \( m(\theta) \) over a compact parameter space \( \Theta \):

\[ \Theta_0 = \{ \theta \in \Theta : m(\theta) \leq 0 \}. \]

A natural estimator of \( \Theta_0 \) is its empirical analog

\[ \hat{\Theta}_0 = \{ \theta \in \Theta : \hat{m}(\theta) \leq 0 \}, \]
where \( \hat{m}(\theta) \) is the estimate of the inequality-generating function. For example, in HJ and M examples, the estimate takes the form

\[
\hat{m}(\theta) = m(\theta, \hat{\gamma}), \quad \hat{\gamma} = \text{vec} \ (\hat{\vartheta}, \hat{\Sigma}).
\]

Our proposals for confidence regions are based on (1) LR-type statistic and (2) Wald-type statistic. The LR-based confidence region is

\[
R_{LR} = \{ \theta \in \Theta : \left[ \sqrt{n} \hat{m}(\theta) / s(\theta) \right]_+^2 \leq \hat{k}(1 - \alpha) \}, \quad (3.2)
\]

where \( s(\theta) \) is the weighting function; ideally, the standard error of \( \hat{m}(\theta) \); and \( \hat{k}(1 - \alpha) \) is a suitable estimate of

\[
k(1 - \alpha) = (1 - \alpha) - \text{quantile of } \mathcal{L}_n,
\]

where

\[
\mathcal{L}_n = \sup_{\theta \in \Theta_0} \left[ \sqrt{n} \hat{m}(\theta) / s(\theta) \right]_+^2 \quad (3.3)
\]

is the LR-type statistic, as in Chernozhukov et al. [2007].

Our Wald-based confidence region is

\[
R_W = \{ \theta \in \Theta : \left[ \sqrt{n} d(\theta, \hat{\Theta}_0) / w(\theta) \right] \leq \hat{k}(1 - \alpha) \}, \quad (3.4)
\]

where \( w(\theta) \) is the weighting function, particular forms of which we will suggest later; and \( \hat{k} \) is a suitable estimate of

\[
k(1 - \alpha) = (1 - \alpha) - \text{quantile of } \mathcal{W}_n,
\]

where \( \mathcal{W}_n \) is the weighted \( W \)-statistic

\[
\mathcal{W}_n = \sup_{\theta \in \Theta_0} \left[ \sqrt{n} d(\theta, \hat{\Theta}_0) / w(\theta) \right]^2. \quad (3.5)
\]
Recall that quantity $d(\theta, \hat{\Theta}_0)$ is the distance of a point $\theta$ to a set $\hat{\Theta}_0$, that is,

$$d(\theta, \hat{\Theta}_0) := \inf_{\theta' \in \hat{\Theta}_0} \|\theta - \theta'\|.$$ 

In the special case, where the weight function is flat, namely $w(\theta) = w$ for all $\theta$, the $W$-statistic $W_n$ becomes the canonical directed Hausdorff distance (Molchanov [1998], Beresteanu and Molinari [2008]):

$$\sqrt{W_n} \propto d(\Theta_0, \hat{\Theta}_0) = \sup_{\theta \in \Theta_0} \inf_{\theta' \in \hat{\Theta}_0} \|\theta - \theta'\|.$$ 

The weighted statistic (3.5) is generally not a distance, but we argue that it provides a very useful extension of the canonical directed Hausdorff distance. In fact, in our empirical example precision weighting dramatically improves the confidence regions.

### 3.2.2 A Basic Limit Theorem for LR and W statistics

In this subsection, we develop a basic result on the limit laws of the LR and $W$ statistics. We will develop this result under the following general regularity conditions:

**R.1** The estimates $\theta \mapsto \hat{m}(\theta)$ of the inequality-generating function $\theta \mapsto m(\theta)$ are asymptotically Gaussian, namely, we have that in the metric space of bounded functions $\ell^\infty(\Theta)$

$$\sqrt{n}(\hat{m}(\theta) - m(\theta)) =_d G(\theta) + o_P(1),$$

where $G(\theta)$ is a Gaussian process with zero mean and a non-degenerate covariance function.

**R.2** Functions $\theta \mapsto \hat{m}(\theta)$ and $\theta \mapsto m(\theta)$ admit continuous gradients $\nabla_{\theta} \hat{m}(\theta)$ and $\nabla_{\theta} m(\theta)$ over the domain $\Theta$, with probability one, where the former is a uniformly consistent estimate of the latter, namely uniformly in $\theta \in \Theta$

$$\nabla_{\theta} \hat{m}(\theta) = \nabla_{\theta} m(\theta) + o_P(1).$$

Moreover, the norm of the gradient $\|\nabla_{\theta} m(\theta)\|$ is bounded away from zero.
R.3 Weighting functions satisfy uniformly in \( \theta \in \Theta \)

\[
s(\theta) = \sigma(\theta) + \alpha_p(1), \quad w(\theta) = \omega(\theta) + \alpha_p(1),
\]

where \( \sigma(\cdot) \geq 0 \) and \( \omega(\cdot) \geq 0 \) are continuous functions bounded away from zero.

In Condition R.1, we require the estimates of the inequality-generating functions to satisfy a uniform central limit theorem. There are plenty of sufficient conditions for this to hold provided by the theory of empirical processes. In our example, this condition will follow from asymptotic normality of the estimates of the mean returns and covariance of returns. In Condition R.2, we require that gradient of the estimate of the inequality-generating function is consistent for the gradient of the inequality-generating function. Moreover, we require that the minimal eigenvalue of \( \nabla \theta m(\theta) \nabla \theta m(\theta)' \) is bounded away from zero, which is an identification condition that allows us to estimate, at a usual speed, the boundary of the set \( \Theta_0 \), which we define as

\[
\partial \Theta_0 := \{ \theta \in \Theta : m(\theta) = 0 \}.
\]

In Condition R.3, we require that the estimates of the weight functions are consistent for the weight functions, which are well-behaved.

Under these conditions we can state the following general result.

**Theorem 1 (Limit Laws of LR and W Statistics).** Under R.1-R.3

\[
\mathcal{L}_n \overset{d}{=} \mathcal{L} + o_p(1), \quad \mathcal{L} = \sup_{\theta \in \partial \Theta_0} \left[ \frac{G(\theta)}{\sigma(\theta)} \right]^2,
\]

\[
\mathcal{W}_n \overset{d}{=} \mathcal{W} + o_p(1), \quad \mathcal{W} = \sup_{\theta \in \partial \Theta_0} \left[ \frac{G(\theta)}{\| \nabla \theta m(\theta) \| \cdot \omega(\theta)} \right]^2,
\]

where both \( \mathcal{W} \) and \( \mathcal{L} \) have distribution functions that are continuous at their \((1 - \alpha)\)-quantiles for \( \alpha < 1/2 \). The two statistics are asymptotically equivalent under the following condition:

\[
\mathcal{W}_n \overset{d}{=} \mathcal{L}_n + o_p(1) \quad \text{if} \quad w(\theta) = \frac{\| \nabla \theta m(\theta) \|}{\sigma(\theta)} \quad \text{for each} \quad \theta \in \Theta.
\]
We see from this theorem that the LR and W statistics converge in law to well-behaved random variables that are continuous transformations of the limit Gaussian process $G(\theta)$. Moreover, we see that under an appropriate choice of the weighting functions, the two statistics are asymptotically equivalent.

For our application to HJ and M sets, the following conditions will be sufficient

**C.1 Estimator of the true parameter value**

The estimator of the true parameter value $\gamma_0$ characterizing the inequality generating function $m(\theta) = m(\theta, \gamma_0)$, where $\gamma_0$ denotes the true parameter value, is such that $\sqrt{n}(\hat{\gamma} - \gamma_0) \rightarrow_d \Omega^{1/2}Z$, $Z = N(0, I_d)$.

**C.2 Gradients**

The gradients $\nabla_\theta m(\theta, \gamma)$ and $\nabla_\gamma m(\theta, \gamma)$ are continuous over the compact parameter space $(\theta, \gamma) \in \Theta \times \Gamma$, where $\Gamma$ is some set that includes an open neighborhood of $\gamma_0$. Moreover, the minimal eigenvalue of $\nabla_\theta m(\theta, \gamma)\nabla_\gamma m(\theta, \gamma)'$ is bounded away from zero over $(\theta, \gamma) \in \Theta \times \Gamma$.

It is straightforward to verify that these conditions hold for the canonical versions of the HJ and M problems.

Under these conditions we immediately conclude that the following approximation is true uniformly in $\theta$, that is, in the metric space of bounded functions $\ell^\infty(\Theta)$:

$$\sqrt{n}(\hat{m}(\theta) - m(\theta)) = \nabla_\gamma m(\theta, \gamma)'\sqrt{n}(\hat{\gamma} - \gamma_0) + o_p(1)$$

$$= d\nabla_\gamma m(\theta, \gamma_0)'\Omega^{1/2}Z + o_p(1),$$

(3.8)

where $\nabla m(\theta, \gamma)$ denotes the gradient with each of its rows evaluated at a value $\tilde{\gamma}$ on the line connecting $\gamma$ and $\gamma_0$, where value $\tilde{\gamma}$ may vary from row to row of the matrix.

Therefore, the limit process in HJ and M examples takes the form:

$$G(\theta) = \nabla_\gamma m(\theta, \gamma_0)'\Omega^{1/2}Z.$$
This will lead us to conclude formally below that conclusions of Theorem 1 hold with

\[
\mathcal{L} = \sup_{\theta \in \mathcal{D}_0} \left[ \frac{\nabla_{\gamma} m(\theta, \gamma) \Omega^{1/2}}{\sigma(\theta)} Z \right]_+^2, \tag{3.11}
\]

\[
\mathcal{W} = \sup_{\theta \in \mathcal{D}_0} \left[ \frac{\nabla_{\gamma} m(\theta, \gamma) \Omega^{1/2}}{\|\nabla_{\theta} m(\theta, \gamma)\| \cdot \omega(\theta)} Z \right]_+^2. \tag{3.12}
\]

A good strategy for choosing the weighting function for LR and W is to choose the studentizing Anderson-Darling weights

\[
\sigma(\theta) = \|\nabla_{\gamma} m(\theta, \gamma_0) \Omega^{1/2}\|, \tag{3.13}
\]

\[
\omega(\theta) = \frac{\|\nabla_{\gamma} m(\theta, \gamma_0) \Omega^{1/2}\|}{\|\nabla_{\theta} m(\theta, \gamma_0)\|}. \tag{3.14}
\]

The natural estimates of these weighting functions are given by the following plug-in estimators:

\[
s(\theta) := \|\nabla_{\gamma} m(\theta, \gamma_0) \widehat{\Omega}^{1/2}\|, \tag{3.15}
\]

\[
w(\theta) := \frac{\|\nabla_{\gamma} m(\theta, \gamma_0) \widehat{\Omega}^{1/2}\|}{\|\nabla_{\theta} m(\theta, \gamma_0)\|}. \tag{3.16}
\]

We formalize the preceding discussion as the following corollary.

**Corollary 1 (Limit Laws of LR and W statistics in HJ and M problems).** Suppose that Conditions C.1-C.2 hold. Then conditions R.1 and R.2 hold with the limit Gaussian process stated in equation (3.10). Furthermore, the plug-in estimates of the weighting functions (3.15) and (3.16) are uniformly consistent for the weighting functions (3.13) and (3.14), so that Condition R.3 holds. Therefore, conclusions of Theorem 1 hold with the limit laws for our statistics given by the laws of random variables stated in equations (3.11) and (3.12).
3.2.3 Basic Validity of the Confidence Regions

In this section we shall suppose that we have suitable estimates of the quantiles of LR and W statistics and will verify basic validity of our confidence regions. In the next section we will provide a construction of such suitable estimates by the means of bootstrap and simulation.

Our result is as follows.

**Theorem 2** (Basic Inferential Validity of Confidence Regions). Suppose that for $\alpha < 1/2$ we have consistent estimates of quantiles of limit statistics $W$ and $L$, namely,

$$\hat{k}(1-\alpha) = k(1-\alpha) + o_p(1), \quad (3.17)$$

where $k(1-\alpha)$ is $(1-\alpha)$-quantile of either $W$ or $L$. Then as the sample size $n$ grows to infinity, confidence regions $R_{LR}$ and $R_W$ cover $\Theta_0$ with probability approaching $1 - \alpha$:

$$\Pr_p[\Theta_0 \subseteq R_{LR}] = \Pr_p[L_n \leq \hat{k}(1-\alpha)] \to \Pr_p[L \leq k(1-\alpha)] = (1-\alpha)(3.18)$$

$$\Pr_p[\Theta_0 \subseteq R_W] = \Pr_p[W_n \leq \hat{k}(1-\alpha)] \to \Pr_p[W \leq k(1-\alpha)] = (1-\alpha)(3.19)$$

The result further applies to HJ and M problems.

**Corollary 2** (Limit Laws of LR and W statistics in HJ and M problems). Suppose that Conditions C.1-C.2 hold and that consistent estimates of quantiles of statistics (3.11) and (3.12) are available. Then conclusions of Theorem 2 apply.
3.2.4 Estimation of Quantiles of LR and W Statistics by Bootstrap and Other Methods

In this section we show how to estimate quantiles of LR and W statistics using bootstrap, simulation, and other re sampling schemes under general conditions. The basic idea is as follows: First, let us take any procedure that consistently estimates the law of our basic Gaussian process \( G \) or a weighted version of this process appearing in the limit expressions. Second, then we can show with some work that we can get consistent estimates of the laws of LR and W statistics, and thus also obtain consistent estimates of their quantiles. It is well-known that there are many procedures for accomplishing the first step, including such common schemes as the bootstrap, simulation, and sub sampling, including both cross-section and time series versions.

In what follows, we will ease the notation by writing our limit statistics as a special case of the following statistic:

\[
S = \sup_{\theta \in \Theta_0} [V(\theta)]_+, \quad V(\theta) = \tau(\theta)G(\theta). \tag{3.20}
\]

Thus, \( S = \mathcal{L} \) for \( \tau(\theta) = 1/s(\theta) \) and \( S = \mathcal{W} \) for \( \tau(\theta) = 1/\|\nabla m(\theta)\| \cdot \omega(\theta) \). We take \( \tau \) to be a continuous function bounded away from zero on the parameter space.

We also need to introduce the following notations and concepts. Our process \( V \) is a random element that takes values in the metric space of continuous functions \( C(\Theta) \) equipped with the uniform metric. The underlying measure space is \( (\Omega, \mathcal{F}) \) and we denote the law of \( V \) under the probability measure \( P \) by the symbol \( Q_V \).

Suppose we have an estimate \( Q_{V^*} \) of the law \( Q_V \) of the Gaussian process \( V \). This estimate \( Q_{V^*} \) is a probability measure generated as follows. Let us fix another measure space \( (\Omega', \mathcal{F}') \) and a probability measure \( P^* \) on this space, then given a random element \( V^* \) on this space taking values in \( C(\Theta) \), we denote its law under \( P^* \) by \( Q_{V^*} \). We thus identify the probability measure \( P^* \) with a data-generating process by which we generate draws or realizations of \( V^* \). This identification allows us to encompass such methods of producing realizations of \( V^* \) as the bootstrap, sub sampling, or other simulation methods. We require that the estimate \( Q_{V^*} \) is
consistent for $Q_V$ in any metric $\rho_K$ metrizing weak convergence, where we can take the metric to be the Kantarovich-Rubinstein metric. Let us mention right away that there are many results that verify this basic consistency condition for various rich forms of processes $V$ and various bootstrap, simulation, and sub sampling schemes for estimating the laws of these processes, as we will discuss in more detail below.

In order to recall the definition of the Kantarovich-Rubinstein metric, let $\theta \mapsto v(\theta)$ be an element of a metric space $(M, d)$, and $Lip(M)$ be a class of Lipschitz functions $\varphi : M \to \mathbb{R}$ that satisfy:

$$|\varphi(v) - \varphi(v')| \leq d(v, v') \wedge 1, \quad |\varphi(v)| \leq 1,$$

The Kantarovich-Rubinstein distance between probability laws $Q$ and $Q'$ is

$$\rho_K(Q, Q'; M) := \sup_{\varphi \in Lip(M)} |E_Q \varphi - E_{Q'} \varphi|.$$ 

As stated earlier, we require that the estimate $Q_{V^*}$ is consistent for $Q_V$ in the metric $\rho_K$, that is

$$\rho_K(Q_{V^*}, Q_V; C(\Theta)) = o_p(1). \quad (3.21)$$

Let $Q_S$ denote the probability law of $S = W$ or $L$, which is in turn induced by the law $Q_V$ of the Gaussian process $V$. We need to define the estimate $Q_{S^*}$ of this law. First, we define the following plug-in estimate of the boundary set $\partial \Theta_0$, which we need to state here:

$$\widehat{\partial \Theta_0} = \{\theta \in \Theta : \widehat{m}(\theta) = 0\}. \quad (3.22)$$

This estimate turns out to be consistent at the usual root-$n$ rate, by the argument like the one given in Chernozhukov et al. [2007]. Then define $Q_{S^*}$ as the law of the following random variable

$$S^* = \sup_{\theta \in \widehat{\partial \Theta_0}} [V^*(\theta)]_+ \quad (3.23)$$

In this definition, we hold the hatted quantities fixed, and the only random element
is \( V^* \) that is drawn according to the law \( Q_{V^*} \).

We will show that the estimated law \( Q_{S^*} \) is consistent for \( Q_S \) in the sense that

\[
\rho_K(Q_{S^*}, Q_S; \mathbb{R}) = o_p(1). \tag{3.24}
\]

Consistency in the Kantarovich-Rubinstein metric in turn implies consistency of the estimates of the distribution function at continuity points, which in turn implies consistency of the estimates of the quantile function.

Equipped with the notations introduced above we can now state our result.

**Theorem 3 (Consistent Estimation of Quantiles)** Suppose Conditions R.1-R.3 hold, and any mechanism, such as bootstrap or other method, is available, which provides a consistent estimate of the law of our limit Gaussian processes \( V \), namely equation (3.21) holds. Then, the estimates of the laws of the limit statistics \( S = \mathcal{W} \) or \( \mathcal{L} \) defined above are consistent in the sense of equation (3.24). As a consequence, we have that the estimates of the quantiles are consistent in the sense of equation (3.17).

We now specialize this result to the HJ and M problems. We begin by recalling that our estimator satisfies

\[
\sqrt{n}(\hat{\gamma} - \gamma) \overset{d}{=} \Omega^{1/2} Z + o_p(1).
\]

Then our limit statistics take the form:

\[
S = \sup_{\theta \in \Theta_0} [V(\theta)]^2, \quad V(\theta) = t(\theta)'Z,
\]

where \( t(\theta) \) is a vector valued weight function, in particular, for \( S = \mathcal{L} \) we have \( t(\theta) = (\nabla_{\gamma} m(\theta, \gamma)'\Omega^{1/2})/\sigma(\theta) \) and for \( S = \mathcal{W} \) we have \( t(\theta) = (\nabla_{\gamma} m(\theta, \gamma)'\Omega^{1/2})/||\nabla_{\theta} m(\theta, \gamma)||\cdot \omega(\theta) \). Here we shall assume that we have a consistent estimate \( Q_{Z^*} \) of the law \( Q_Z \).
of $Z$, in the sense that,

$$ \rho_K(Q_{Z^*}, Q_Z) = o_p(1). \quad (3.25) $$

There are many methods that provide such consistent estimates of the laws. Bootstrap is known to be valid for various estimation methods (van der Vaart and Wellner [1996]); simulation method that simply draws $Z \sim N(0, I)$ is another valid method; and subsampling is another rather general method (Politis and Romano [1994]). Next, the estimate $Q_{V^*}$ of the law $Q_{V^*}$ is then defined as:

$$ V^*(\theta) = \hat{t}(\theta)' Z^*, \quad (3.26) $$

where $\hat{t}(\theta)$ is a vector valued weighting function that is uniformly consistent for the weighting function $t(\theta)$. In this definition we hold the hatted quantity fixed, and the only random element is $Z^*$ that is drawn according to the law $Q_{Z^*}$. Then, we define the random variable

$$ S^* = \sup_{\theta \in \Theta_0} [V^*(\theta)]^2, $$

and use its law $Q_{S^*}$ to estimate the law $Q_S$.

We can now state the following corollary.

\textbf{Corollary 3 (Consistent Estimation of Quantiles in HJ and M problems)} Suppose Conditions C.1-C.2 hold, and any mechanism, such as bootstrap or other method, that provides a consistent estimate of the law of $Z$ is available, namely equation (3.25) holds. Then, this provides us with a consistent estimate of the law of our limit Gaussian process $G$, namely equation (3.21) holds. Then, all of the conclusions of Theorem 3 hold.
3.3 Empirical Example

As an empirical example we use HJ bounds which are widely used in testing asset pricing models. In order to keep results comparable, the sample used in this section is very similar to data used in Hansen and Jagannathan [1991]. The two asset series used are annual treasury bond returns and annual NYSE value-weighted dividend included returns. These nominal returns are converted to real returns by using implicit price deflator based on personal consumption expenditures as in Hansen and Jagannathan [1991]. Asset returns are from CRSP, and the implicit price deflator is available from St. Louis Fed and based on National Income and Product Accounts of United States. We use data for the time period 1959-2006 (inclusive).

Figures can be found in Subsection 3.5.1. Figure 3-1 simply traces out the mean-standard deviation pairs which satisfy

\[ m(\theta, \hat{\gamma}) = 0 \]

where \( \hat{\gamma} \) is estimated using sample moments.

Figure 3-2 represents the uncertainty caused by the estimation of \( \gamma \). To estimate the distribution of \( \hat{\gamma} \) bootstrap method is used. Observations are drawn with replacement from the bivariate time series of stock and bond returns. 100 bootstraps result in 100 \( \hat{\gamma} \). The resulting HJ bounds are included in the figure.

In Figure 3-3 in addition to the bootstrapped curves 90% confidence region based on LR statistic is presented. LR based confidence region covers most of the bootstrap draws below the HJ bounds as expected. An attractive outcome of using this method is that the resulting region does not include any unnecessary areas that is not covered by bootstrap draws.

Figure 3-4 plots 90% confidence region based on unweighted LR statistic. Comparison of Figure 3-3 and Figure 3-4 reveals that precision weighting plays a very important role in delivering good confidence sets. Without precision weighting LR statistic delivers a confidence region that includes unlikely regions in the parameter space where standard deviation of the discount factor is zero. On the other hand pre-
cision weighted LR based confidence region is invariant to parameter transformations, for example, changes in units of measurement. This invariance to parameter transformations is the key property of a statistic to deliver desirable confidence regions that does not cover unnecessary areas.

Figure 3-5 plots confidence region based on Wald-based statistic with no precision weighting. This is identical to the confidence region based on Hausdorff distance. Similar to Figure 3-4 this region covers a large area of the parameter space where no bootstrap draws appear. This picture reveals a key weakness of using an unweighted Wald-based statistic or Hausdorff distance to construct confidence regions. These methods are not invariant to parameter transformations which results in confidence regions with undesirable qualities that cover unnecessary areas in the parameter space.

The problem in Figure 3-4 and Figure 3-5 are of similar nature. In both of these cases the statistics underlying the confidence regions are not invariant to parameter transformations therefore when drawing confidence regions uncertainty in one part of the plot is assumed to be identical to uncertainty in other parts of the plot. However a quick look at the Figure 3-2 reveals that uncertainty regarding the location of the HJ bound varies for a given mean or standard deviation of the stochastic discount factor.

Figure 3-6 plots the confidence region based on weighted Wald statistic. Weighting fixes the problem and generates a statistic that is invariant to parameter transformations. The resulting confidence set looks very similar to weighted LR based confidence set in Figure 3-3 as it covers most of the bootstrap draws below the HJ bounds and does not include unnecessary regions in the parameter space.

3.4 Conclusion

In this paper we provided various inferential procedures for inference on sets that solve a system of inequalities. These procedures are useful for inference on Hansen-Jagannathan mean-variance sets of admissible stochastic discount factors and Markowitz mean-variance sets of admissible portfolios.
3.5 Appendix

3.5.1 Figures

Figure 3-1: Estimated HJ Bounds
Figure 3-2: Estimated HJ Bounds and Bootstrap Draws
Figure 3-3: 90% Confidence Region using LR Statistic
Figure 3-4: 90% Confidence Region using Unweighted LR Statistic
Figure 3-5: 90% Confidence Region using Unweighted W Statistic (H-Distance)
Figure 3-6: 90% Confidence Region using Weighted W Statistic
3.5.2 Proofs

First, we will prove a simple lemma which will be used to justify the local approximation for the Wald statistic:

**Lemma 1.** Suppose R.1 and R.2 hold, and let \( \theta_n \) be a sequence such that \( \theta_n \to \theta^* \in \partial \Theta_0 \). Then \( \bar{\theta}_n \equiv \arg\min_{\theta : m(\theta) = 0} \| \theta_n - \theta \|^2 \) satisfies \( \bar{\theta}_n \to \theta^* \).

**Proof of Lemma 1.** Since by assumption R.2, the gradient of \( \hat{m}(\theta) \) is bounded away from zero, by the implicit function theorem, the set \( \{ \theta : m(\theta) = 0 \} \) is locally approximated by a plane, and we can define

\[
\bar{\theta}_n = \arg \min_{\theta : m(\theta) = 0} \| \theta^* - \theta \|^2 = \theta^* + \nabla \theta m(\theta^*) (\nabla \theta m(\theta^*) \nabla \theta m(\theta^*))^{-1} (\hat{m}(\theta^*) - m(\theta^*)) + o_p(1)
\]

By R.1 and R.2, \( \bar{\theta}_n - \theta^* = o_p(1) \) so that by the triangle inequality

\[
\| \bar{\theta}_n - \theta^* \| \leq \| \bar{\theta}_n - \theta_n \| + \| \theta_n - \theta^* \| \\
\leq \| \theta_n - \theta^* \| + \| \theta_n - \bar{\theta}_n \| \\
\leq 2\| \theta_n - \theta^* \| + \| \theta^* - \bar{\theta}_n \| = o_p(1)
\]

since \( \theta_n \to \theta^* \) and \( \bar{\theta}_n - \theta^* = o_p(1) \).

**Proof of Theorem 1**

**PART 1.** (Limit law of \( \mathcal{L}_n \)) Let \( G_n = \sqrt{n}(\hat{m} - m) \). Then

\[
\mathcal{L}_n = \sup_{\theta \in \Theta_0} \left[ \frac{\sqrt{n}m(\theta)/s(\theta)}{s(\theta)} \right]_+^2 = \sup_{\theta \in \Theta_0} \left[ \frac{(G_n(\theta) + \sqrt{n}m(\theta))/s(\theta)}{s(\theta)} \right]_+^2 \\
= d \sup_{\theta \in \Theta_0} \left[ \frac{(G(\theta) + \sqrt{n}m(\theta))/\sigma(\theta) + o_p(1)}{\sigma(\theta)} \right]_+^2 \\
= \sup_{\theta \in \partial \Theta_0} \left[ \frac{(G(\theta) + \sqrt{n}m(\theta))/\sigma(\theta) + o_p(1)}{\sigma(\theta)} \right]_+^2
\]

The steps, apart from the last, immediately follow from Conditions R.1 and R.3. The last step follows from the argument given below. Indeed, take any sequence \( \theta_n \in \Theta_0 \).
such that

$$\sup_{\theta \in \Theta_0} \left[ \frac{(G(\theta) + \sqrt{nm(\theta)})/\sigma(\theta) + o_p(1)}{\sqrt{nm(\theta)}} \right] = \sup_{\theta \in \Theta_0} \left[ \frac{(G(\theta) + \sqrt{nm(\theta)})/\sigma(\theta)}{\sqrt{nm(\theta)}} \right].$$

In order for this to occur we need to have that

$$\sqrt{nm(\theta_n)/\sigma(\theta_n)} = O_p(1),$$

which is only possible in view of condition R.2 if, for some stochastically bounded sequence of positive random variables $C_n = O_p(1),$

$$\sqrt{n}d(\theta_n, \partial \Theta_0) \leq C_n.$$

Therefore we conclude that

$$\sup_{\theta \in \Theta_0} \left[ \frac{(G(\theta) + \sqrt{nm(\theta)})/\sigma(\theta) + o_p(1)}{\sqrt{nm(\theta)}} \right] = \sup_{\theta \in \Theta_0} \left[ \frac{(G(\theta) + \sqrt{nm(\theta)})/\sigma(\theta)}{\sqrt{nm(\theta)}} \right].$$

Using stochastic equicontinuity of $G$ and continuity of $\sigma,$ the last quantity is further approximated by

$$\sup_{\theta \in \Theta_0, \theta + \lambda/\sqrt{nm(\theta)} \in \Theta_0, ||\lambda|| \leq C_n} \left[ \frac{(G(\theta) + \sqrt{nm(\theta) + \lambda}/\sqrt{n})/\sigma(\theta + \lambda/\sqrt{n}) + o_p(1)}{\sqrt{nm(\theta) + \lambda}/\sqrt{n}} \right].$$

Because $\sqrt{nm(\theta + \lambda)} \leq 0$ and $m(\theta) = 0$ for $\theta \in \partial \Theta_0$ and $\theta + \lambda/\sqrt{n} \in \Theta_0,$ we conclude that the last quantity is necessarily equal to $\sup_{\theta \in \Theta_0} (G(\theta)/\sigma(\theta))^2,$ yielding the conclusion we needed.

**PART 2.**  (*Limit Law of $\mathcal{W}_n$*). We will begin by justifying the approximation holding with probability going to one

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, \widehat{\Theta}_0) = \sup_{\Theta_n} \sqrt{n}d(\theta, \widehat{\Theta}_0). \quad (3.27)$$
where

$$\Theta_n = \{\theta \in \Theta_0 : \sqrt{n}d(\theta, \partial \Theta_0) \leq C_n\}$$

where $C_n$ is some stochastically bounded sequence of positive random variables, $C_n = \text{O}_p(1)$. Note that right hand side is less than or equal to the left hand side in general, so we only need to show that the right hand side can not be less. Indeed, let $\theta_n$ be any sequence such that

$$\sup_{\theta \in \Theta_0} \sqrt{n}d(\theta, \hat{\Theta}_0) = \sqrt{n}d(\theta_n, \hat{\Theta}_0).$$

If $\hat{m}(\theta_n) \leq 0$, then $d(\theta_n, \hat{\Theta}_0) = 0$, and the claim follows trivially since the right hand side of (3.27) is non-negative and is less than or equal to the left hand side of (3.27). If $\hat{m}(\theta_n) > 0$, then $d(\theta_n, \hat{\Theta}_0) > 0$, but for this and for $\theta_n \in \Theta_0$ to take place we must have that $0 < \hat{m}(\theta_n) = \text{O}_p(1/\sqrt{n})$, which by Condition R.2 implies that $d(\theta_n, \hat{\Theta}_0) = \text{O}_p(1/\sqrt{n})$.

In the discussion the quantity $\theta^*(\theta)$ as follows

$$\theta^*(\theta) \in \arg \min_{\theta' \in \Theta_0} \|\theta - \theta'\|^2.$$ 

The argmin set $\theta^*(\theta)$ is a singleton simultaneously for all $\theta \in \Theta_n$, provided $n$ is sufficiently large. This follows from condition R.2 imposed on the gradient $\nabla_{\theta} m$. Moreover, by examining the optimality condition we can conclude that we must have that for $\theta \in \Theta_n$

$$(I - \nabla_{\theta} m(\theta)(\nabla_{\theta} m(\theta)')(\nabla_{\theta} m(\theta))^{-1}\nabla_{\theta} m(\theta)')(\theta - \theta^*) = \text{O}_p(1) \quad (3.28)$$

The projection of $\theta \in \Theta$ onto the set $\hat{\Theta} := \{\theta \in \Theta : \hat{m}(\theta) \leq 0\}$ is given by

$$\tilde{\theta}(\theta) = \arg \min_{\theta, \hat{m}(\theta) \leq 0} \|\theta - \theta'\|^2.$$
If $\hat{m}(\theta) \leq 0$, then $\tilde{\theta}(\theta) = \theta$. If $\hat{m}(\theta) > 0$, then $\tilde{\theta}(\theta) = \bar{\theta}(\theta)$, where

$$
\bar{\theta}(\theta) = \arg \min_{\theta \cdot \hat{m}(\theta') = 0} \| \theta - \theta' \|^2.
$$

In what follows we will suppress the indexing by $\theta$ in order to ease the notation, but it should be understood that we will make all the claims uniformly in $\theta \in \Theta_n$. For each $\theta$, the Lagrangian for this problem is $\| \theta - \theta' \|^2 + 2\hat{m}(\theta')\lambda$. Therefore, the quantity $\tilde{\theta}(\theta)$ can be take to be an interior solution of the saddle-point problem

$$(\tilde{\theta} - \theta) + \nabla_\theta \hat{m}(\tilde{\theta})\lambda = 0$$
$$\hat{m}(\tilde{\theta}) = 0$$

The corner solutions do not contribute to the asymptotic behavior of $\mathcal{W}_n$, and thus can be ignored. A formal justification for this will be presented in future versions of this work. Using mean-value expansion we obtain

$$(\tilde{\theta} - \theta) + \nabla_\theta \hat{m}(\tilde{\theta})\lambda = 0$$
$$m(\theta^*) + \nabla_\theta m(\tilde{\theta})(\tilde{\theta} - \theta^*) + \hat{m}(\tilde{\theta}) - m(\tilde{\theta}) = 0$$

Since $\nabla_\theta \hat{m}(\tilde{\theta}) = \nabla_\theta m(\theta) + o_p(1)$ and $\nabla_\theta m(\tilde{\theta}) = \nabla_\theta m(\theta) + o_p(1)$ uniformly in $\theta \in \Theta$, solving for $(\tilde{\theta} - \theta)$ we obtain

$$\tilde{\theta} - \theta^* = [\nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1} + o_p(1)](\hat{m}(\tilde{\theta}) - m(\tilde{\theta}))$$
$$+ (I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}\nabla_\theta m(\theta)' + o_p(1))(\theta - \theta^*)$$

Using that $\sqrt{n}(\hat{m}(\theta) - m(\theta)) = d G(\theta) + o_p(1)$, we obtain

$$\sqrt{n}(\tilde{\theta} - \theta^*) = d \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}G(\theta)$$
$$+ (I - \nabla_\theta m(\theta)(\nabla_\theta m(\theta)'\nabla_\theta m(\theta))^{-1}\nabla_\theta m(\theta)')(\theta - \theta^*)$$

Furthermore, by $\theta \in \Theta_n$ and by the approximate orthogonality condition (3.28) we
further have that $(I - \nabla_{\theta} m(\theta)(\nabla_{\theta} m(\theta)'\nabla_{\theta} m(\theta))^{-1}\nabla_{\theta} m(\theta)'(\theta - \theta^*)) = o_p(1)$, so that

$$\sqrt{n}(\hat{\theta} - \theta^*) = d \nabla_{\theta} m(\theta)(\nabla_{\theta} m(\theta)'\nabla_{\theta} m(\theta))^{-1}G(\theta) + o_p(1).$$

We next approximate $1(\hat{m}(\theta) > 0)$ using that

$$\sqrt{n}\hat{m}(\theta) = \sqrt{n}\hat{m}(\bar{\theta}) + \nabla_{\theta} m(\bar{\theta})\sqrt{n}(\theta - \bar{\theta}) = \nabla m(\theta)'\sqrt{n}(\theta - \bar{\theta}) + o_p(1),$$

for an intermediate value $\bar{\theta}$, where we used that $\hat{m}(\bar{\theta}) = 0$.

Thus, uniformly in $\theta \in \Theta_n$ we have that

$$\sqrt{n}d(\theta, \hat{\Theta}_n) = \|\bar{\theta} - \theta\|^21\{\nabla m(\theta)\sqrt{n}(\theta - \bar{\theta}) > 0 + o_p(1)\}$$

$$= |\nabla_{\theta} m(\theta)'\nabla_{\theta} m(\theta))^{-1/2}G(\theta)|1\{G(\theta) > 0 + o_p(1)\}$$

$$= [\|\nabla_{\theta} m(\theta)\|^{-1}G(\theta) + o_p(1)]_+$$

Therefore, given the initial approximation (3.27) we obtain that

$$\mathcal{W}_n = d \sup_{\theta \in \bar{\Theta}_n} [\|\nabla_{\theta} m(\theta)\|^{-1}G(\theta)]_+ + o_p(1).$$

PART 3. (Continuity of the Limit Distributions). The continuity of the distribution function $\mathcal{L}$ on $(0, \infty)$ follows from the Davydov et al. [1998] and from the assumption that the covariance function of $G$ is non-degenerate. Probability that $\mathcal{L}$ is greater than zero is equal to the probability that $\max_j \sup_{\theta \in \Theta} G_j(\theta) > 0$ which is greater than the probability that $G_j'(\theta') > 0$ for some fixed $j'$ and $\theta'$, but the latter is equal to $1/2$. Therefore the claim follows. The claim of continuity of the distribution function of $\mathcal{W}$ on $(0, \infty)$ follows similarly. \qed
Proof of Corollary 1

This corollary immediately follows from the assumed conditions and from the comments given in the main text preceding the statement of Corollary 1. □

Proof of Theorem 2

We have that $Pr_{p}[\Theta_0 \subseteq R_{LR}] = Pr_{p}[\mathcal{L}_n \leq \hat{k}(1 - \alpha)]$ by the construction of the confidence region. We then have that for any $\alpha < 1/2$ that $k(1 - \alpha)$ is a continuity point of the distribution function of $\mathcal{L}$, so that for any sufficiently small $\epsilon$

$$Pr_{p}[\mathcal{L}_n \leq \hat{k}(1 - \alpha)] \leq Pr_{p}[\mathcal{L}_n \leq k(1 - \alpha) + \epsilon] \rightarrow Pr_{p}[\mathcal{L} \leq k(1 - \alpha) + \epsilon]$$

$$Pr_{p}[\mathcal{L}_n \leq \hat{k}(1 - \alpha)] \geq Pr_{p}[\mathcal{L}_n \leq k(1 - \alpha) - \epsilon] \rightarrow Pr_{p}[\mathcal{L} \leq k(1 - \alpha) - \epsilon]$$

Since we can set $\epsilon$ as small as we like and $k(1 - \alpha)$ is a continuity point of the distribution function of $\mathcal{L}$, we have that

$$Pr_{p}[\mathcal{L}_n \leq \hat{k}(1 - \alpha)] \rightarrow Pr_{p}[\mathcal{L} \leq k(1 - \alpha)] = (1 - \alpha).$$

We can conclude similarly for the $W$-statistic $W_n$. □.

Proof of Corollary 2

This corollary immediately follows from the assumed conditions and Corollary 1. □

Proof of Theorem 3

We have that

$$E_{P^*}[\varphi(V^*)] - E_{P}[\varphi(V)] = o_p(1) \text{ uniformly in } \varphi \in Lip(C(\Theta)).$$

This implies that

$$E_{P^*}[\varphi([V^*]_+)] - E_{P}[\varphi([V]_+)] = o_p(1) \text{ uniformly in } \varphi \in Lip(C(\Theta)), $$

113
since the composition $\varphi \circ [\cdot]_+ \in Lip(C(\Theta))$ for $\varphi \in Lip(C(\Theta))$. This further implies that

$$E_{P^*}[\varphi'(\sup_{R_n}[V^*]+)] - E_P[\varphi'(\sup_{R_n}[V]+)] = o_p(1) \text{ uniformly in } \varphi' \in Lip(\mathbb{R}),$$

since the composition $\varphi'(\sup_{R_n}[\cdot]_+) \in Lip(C(\Theta))$ for $\varphi' \in Lip(\mathbb{R})$ and $R_n$ denoting any sequence of closed non-empty subsets in $\Theta$. We have that $\overline{\partial\Theta}_0$ converges to $\partial\Theta_0$ in the Hausdorff distance, so that

$$|E_P[\varphi'(\sup_{\overline{\partial\Theta}_0}[V]+) - \varphi'(\sup_{\partial\Theta_0}[V]+)]|$$

$$\leq E[|\sup_{\overline{\partial\Theta}_0}[V]+ - \sup_{\partial\Theta_0}[V]+| \wedge 1] = o_p(1) \text{ uniformly in } \varphi' \in Lip(\mathbb{R}),$$

since $\sup_{\overline{\partial\Theta}_0}[V]+ - \sup_{\partial\Theta_0}[V] = o_p(1)$ by stochastic equicontinuity of the process $V$. Since metric $\rho_K$ is a proper metric that satisfies the triangle inequality, we have shown that

$$\rho_K(Q_{S^*}, Q_S) = o_p(1).$$

Next, we note that the convergence $\rho_K(Q_{S_n}, Q_S) = o(1)$, for any sequence of laws $Q_{S_n}$ of a sequence of random variables $S_n$ defined on probability space $(\Omega', \mathcal{F}', P_n)$ implies the convergence of the distribution function

$$Pr_{Q_{S_n}}[S_n \leq s] = Pr_{Q_S}[S \leq s] + o(1)$$

at each continuity point $(0, \infty)$ of the mapping $s \mapsto Pr[S \leq s]$ and also convergence of quantile functions

$$\inf\{s : Pr_{Q_{S_n}}[S_n \leq s] \geq p\} = \inf\{s : Pr_{Q_S}[S \leq s] \geq p\} + o(1)$$

at each continuity point $p$ of the mapping $s \mapsto \inf\{s : Pr_{Q_S}[S \leq s] \geq p\}$. Recall from Theorem 1 that the set of continuity points necessarily includes the region $(0, 1/2)$.

By the Extended Continuous Mapping Theorem we conclude that since $\rho_K(Q_{S^*}, Q_S) = $
$o_p(1)$, for any sequence of laws $Q_s^*$ of random variable $S^*$ defined on probability space $(\Omega', \mathcal{F}', P^*)$, we obtain the convergence in probability of the distribution function

$$Pr_{Q_{s^*}}[S^* \leq s] = Pr_{Q_s}[S \leq s] + o_p(1)$$

at each continuity point $(0, \infty)$ of the mapping $s \mapsto Pr[S \leq s]$ and also convergence in probability of the quantile functions

$$\inf\{s : Pr_{Q_{s^*}}[S^* \leq s] \geq p\} = \inf\{s : Pr_{Q_s}[S \leq s] \geq p\} + o_p(1)$$

at each continuity point $p$ of the mapping $s \mapsto \inf\{s : Pr_{Q_s}[S \leq s] \geq p\}$. $\square$

**Proof of Corollary 3**

In order to prove this corollary it suffices to show that

$$\rho_K(Q_{\hat{v}Z^*}, Q_v Z; C(\Theta)) = o_p(1).$$

Without loss of generality we can take $\sup \|\hat{t}\| \leq 1$ and $\sup \|t\| \leq 1$. The claim will follow from

$$\rho_K(Q_{\hat{v}Z^*}, Q_v Z; C(\Theta)) \leq \rho_K(Q_{\hat{v}Z^*}, Q_{\hat{v}Z}; C(\Theta)) + \rho_K(Q_{\hat{v}Z}, Q_v Z; C(\Theta)) = o_p(1).$$

That $\rho_K(Q_{\hat{v}Z^*}, Q_v Z; C(\Theta)) = o_p(1)$ follows immediately from $\rho_K(Q_{Z^*}, Q_Z) = o_p(1)$ and from the mapping $\varphi(\hat{t}') \in Lip(\mathbb{R}^k)$ (indeed, $|\varphi(\hat{t}'Z) - \varphi(t'Z)| \leq \sup |\hat{t}'(z' - z')| \wedge 1 \leq \sup \|t\| \sup \|z - z'\| \wedge 1 \leq [\sup \|t\| \|z - z'\| \wedge 1]$. That $\rho_K(Q_{\hat{v}Z}, Q_v Z; C(\Theta)) = o_p(1)$ follows because uniformly in $\varphi \in Lip(C(\Theta))$

$$|E[\varphi(\hat{t}'Z)] - \varphi(t'Z)| \leq E[\sup |(\hat{t} - t)'Z| \wedge 1] \leq E[\sup \|\hat{t} - t\| |Z| \wedge 1] = o_p(1).$$

$\square$
Bibliography


