Essays on Exchange Rates, and Consumption

by

Pierre-Olivier Gourinchas

Ingénieur, Ecole Nationale des Ponts et Chaussées, Paris, France (1993)
DEA EHESS, Paris, France (1991)
Ingénieur, Ecole Polytechnique, Palaiseau, France (1990)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1996

© 1996 Pierre Olivier Gourinchas. All rights reserved.

The author hereby grants to Massachusetts Institute of Technology permission to
reproduce and
to distribute copies of this thesis document in whole or in part.

Signature of Author:

Department of Economics
May 15 1996

Certified by:

Olivier J. Blanchard
Professor of Economics
Thesis Supervisor

Certified by:

Rudiger Dornbusch
Ford Professor of International Economics
Thesis Supervisor

Accepted by:

Richard S. Eckaus
Chairperson, Departmental Committee on Graduate Students
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Essays on Exchange Rates, and Consumption

by

Pierre-Olivier Gourinchas

Submitted to the Department of Economics
on May 15 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Chapter 2, with Aaron Tornell: Interest rate expectations are essential for exchange rate determination. Using a unique Survey data set on interest rates forecasts from 1986 to 1995 for G7 countries, we find that interest rate shocks were significantly more persistent in sample than expected by the market. This is consistent with Frankel and Froot (1989)'s finding that changes in the forward rate reflect changes in exchange rate expectations. We then present a model of nominal exchange rate determination that rationalizes the forward discount puzzle and exhibits the delayed overshooting pattern found by Eichenbaum and Evans (1995): following a monetary expansion that reduces the domestic interest rate, there is a gradual depreciation of the exchange rate followed by a gradual appreciation several months later. Delayed overshooting results from (a) the interaction of learning about the current state of affairs, and the intrinsic dynamic response of interest rates to monetary shocks and (b) the discrepancy between the actual distribution of shocks in sample and its expectation by market participants. This discrepancy is consistent with rational expectations if either (a) there is a small sample or Peso problem or (b) the true structure of the economy evolves over time and agents are learning with some delay.

Chapter 3: Target zone models of exchange rate imply a non-linear relationship between the fundamental and the exchange rate. Measurement error in constructing the fundamental imply that standard estimation is inconsistent. This paper develops a Non Parametric Instrumental Variable Estimation method to test for exchange rate non-linearities in the presence of measurement error. Non-parametric methods allow consistent instrumental estimation by relaxing the functional specification. For the exchange rate process, I argue that the excess return differential, the excess of stock returns over bonds in the domestic versus the foreign country, constitutes a valid instrument. Applying this procedure to EMS exchange rate data, I show that substantial non-linearities are present, conforming partially with the simplest target zone theory. This approach provides a versatile estimation method that can potentially solve measurement error problems in
general non-linear specifications for a variety of economic situations.

Chapter 4: This paper analyzes the case for fiscal federal transfers in a Monetary Union. Looking at the labor market structure, it emphasizes the incentive effect of any federal transfer scheme insuring workers against bad draws. When the wage negotiation process occurs at the national level and the federal government has incomplete information on the bargaining process, workers have an incentive to ask ex-ante for higher wages. This may negatively affect the macroeconomic performance in the federation. The First Best solution consists in shifting the wage bargaining process from the national to the federal level. Decentralization of fiscal policy would solve the incentive problem. However, looking at the fiscal federalism issue, we show that it is always optimal to keep federal transfers. Moreover, decentralized policies are only effective when access to financial markets is imperfect. Thus the paper makes a strong case in favor of centralization.

Chapter 5, with Jonathan Parker: This paper employs a synthetic cohort technique and Consumer Expenditure Survey data to construct average age-profiles of consumption and income over the working lives of typical households across different education and occupation groups. Even after controlling for family and cohort effects, typical consumption profiles are not flat, and seem to track income at young ages. Using these profiles, we estimate a structural model of optimal life-cycle consumption expenditures in the presence of realistic income uncertainty. The model fits the profiles quite well. In addition to providing tight estimates of the discount rate and risk aversion, we find that consumer behavior changes strikingly over the life-cycle. Young consumers behave as "buffer-stock" agents. Around age 43, the typical household starts accumulating liquid assets for retirement and its behavior mimics more closely that of a certainty equivalent consumer. This change in behavior is mostly driven by the life-cycle profile of expected income. Our methodology provides a natural decomposition of saving into its precautionary and retirement components.

Thesis Supervisor: Olivier J. Blanchard
Title: Professor of Economics

Thesis Supervisor: Rudiger Dornbusch
Title: Ford Professor of International Economics
To my beloved wife, Marion, and daughter, Julie,

They are the salt of the earth.
BIOGRAPHICAL NOTE

Pierre-Olivier Gourinchas was born in Montpellier, France on March 30, 1968. He finished his Baccalauréat degree in 1985 at the Lycée Mas de T esse and went to the Lycée Louis-le-Grand to study maths and physics for two years. In 1990, he graduated from Ecole Polytechnique, where his interest in macroeconomics was first stimulated by a 3 months training course at the research center of the Caisse des Dépots et Consignations in Paris. He then completed simultaneously a Diplôme d'Études Appliquées in Economics at the Ecole des Hautes Études en Sciences Sociales - DELTA and the Ecole Nationale des Ponts et Chaussées in Paris.

After a one year working experience at the research department of the Banque Indosuez, he left in 1992 to attend the Massachusetts Institute of Technology in the pursuit of his doctoral degree in economics. Since 1993, the author is a member of the Centre d'Études et de la Recherche de l'Activité Socio-économique (CERAS), in Paris. His fields of specialization are macroeconomics and international finance, with minors in econometrics and theory. While at MIT, he has worked as a research assistant for Olivier Blanchard and Ricardo Caballero, has been a teaching assistant for Ricardo Caballero, and has lectured many students as the Instructor for MIT's undergraduate International Economy Course.
Acknowledgements

A thesis takes time. Time to learn, time to think, and time to interact. MIT is an exceptional place in all three respects. There, I received an outstanding education, one that stimulates intellectual curiosity and establishes the solid foundations of future research. In the apprentice economists, MIT faculty already see the germ of future colleagues. This attention to students builds confidence and teaches us the values of a modern humanism. Nowhere have I encountered such a collegial place, where doors are wide open, students welcome, and intellectual activity so intense. Let me here thank collectively the MIT Economics Department for being itself, and offering so much.

A special thank goes to my advisors, Olivier Blanchard and Rudiger Dornbusch. Olivier has been a model for me, not simply because he is French! Throughout the years, he has given me extremely valuable support and guidance. He knows how to reorient my ideas at critical junctures and push me forward. His sharp mind always detects the flaw of an argument many steps ahead, as well as seeing the way to resolve it. Rudi taught me more than economics. With him I learned how to be an economist. He always offered penetrating advices at the early stages of a my projects. His openness to students as well as his immense knowledge are unmatched. The International Breakfasts were the place and time for Rudi’s students to get insights and comments. They will remain high in our collective memory.

I owe an immense debt to Ricardo Caballero. By and far, Ricardo had all the at-
tributes of an advisor, except, due to MIT regulation, the ultimate signature. He was always generous with his time and forced me to focus my research more precisely. He read time and again the drafts of papers I gave him. Having been both his teaching and research assistant, I learned immensely from him. His understanding of economics fascinates me and I wish to someday be as talented as he is.

Aaron Tornell was visiting MIT during my second year. I was immediately attracted by his enlightening intuitions. His topics course in International Finance, taught at Harvard, was exceptional. He brought his students to the very frontier of the field. We started a collaboration that lasts to this day and, with time, has become a friendship, filled with exciting discussions. I am looking forward to our next joint research projects.

Many other faculty members also deserve thanks. Daron Acemoglu and I had highly stimulating exchanges. Daron's encyclopedic knowledge of macroeconomics, labor, theory, and development is breathtaking. Julio Rotemberg, Jaume Ventura, Whitney Newey and Andrew Bernard all gave valuable advice during the course of my thesis.

Jacques-François Thisse at CERAS, Paris, provided unflagging support for my research. Charles Wyplosz gave the impetus to my thesis, and beyond, to my research career. As an advisor in France, he taught me the ropes. It is he also, who convinced me of the merits of MIT. He was right, as always.

Without the generous financial assistance from the Ministère de l'Equipement in France, this thesis would not have been possible. I also acknowledge financial support from MIT economics department, the World Economy Laboratory and the Fulbright Program.

The strength of the experience, of course, resides partly in shared, and ultimately transcended, difficulties. MIT is unique for the quality of students it attracts. Onboard the same ship for four years, many of us became friends along the way.

Jonathan Parker and I worked together on problem sets and, by the end of our second year started the project that eventually became chapter 5 of the present thesis.
In two years of intense collaboration, Jonathan and I became close friends. His talents in economics are only matched by his wit. Some of us do remember some of his jokes!

The European connection (Joao Ejarque, Giuseppe Moscarini, Marco Ottaviani, Paola Ramada, Angel Serrat and Karl Whelan) was, of course, always present. Patrizia Canziani’s good spirits warmed our lives. Robert Shimer introduced us to American practice of Thanksgiving. Rodrigo Valdés is the man to beat on the Charles river, where we both went racing for hours, trying to forget the stress of research.

MIT would not function as a department without the help of the support and administrative staff. Many thanks to John Arditi, Theresa Benevento, Gary King, Maureen Maguire and Katherin Swan. They have helped me more than once get ahead of everybody at the copy machine.

Lastly, I want to thank my wife Marion, and my daughter Julie. Julie was born in the US and I remember the day my “maternity beeper” went on during Peter Diamond’s class on general equilibrium. Marion and Julie have lost the count of evenings, week-ends or even nights I spent in the department. Without them, all of this would not have had much sense.
# Contents

Acknowledgements ........................................... 8

1 Introduction ............................................... 23

2 Exchange Rate Dynamics and Learning, joint with Aaron Tornell ....... 31

2.1 Introduction ............................................ 31

2.2 Stylized Facts About Interest Rates Differentials .................... 38

2.2.1 Modelling the Interest Rate Differential ...................... 38

2.2.2 Survey Data .......................................... 42

2.2.3 Maximum Likelihood Estimation of the Interest Rate Process .... 45

2.2.4 Estimating the Market filter ................................ 51

2.3 A Representative Agent Model ................................ 56

2.3.1 The Learning Problem .................................. 58

2.3.2 Market Equilibrium ..................................... 61

2.3.3 Conditional Delayed Overshooting .......................... 63

2.3.4 Unconditional Delayed Overshooting ......................... 70

2.3.5 Predictable Excess Returns ................................ 73

2.4 An Extended Model ....................................... 74

2.4.1 The Learning Problem ................................... 76

2.4.2 The Equilibrium Price Function ............................. 77
2.4.3 Predictable excess returns ....................................... 79
2.4.4 Maximum Likelihood Estimation of the Interest Rate Process ... 80
2.5 Conclusion ............................................................... 83

A Appendix ........................................................................... 85

a Delayed Overshooting ...................................................... 85
b Empirical Results on Interest Rate Differentials ................... 86
  b.1 Kalman Filter Estimation ............................................. 88
c Proofs. .............................................................................. 89

3 Non-Linearities in Exchange Rates: A Non-Parametric Instrumental
  Variable Approach ......................................................... 113

1 Introduction ..................................................................... 113

2 Exchange rate target zones, a brief review of the theory and empirical
  results. ............................................................................. 116
  2.1 The prototype model; .................................................. 116
  2.2 Extensions; .............................................................. 119
  2.3 The empirical literature. .............................................. 121

3 An agnostic approach using Non-Parametric Instrumental Variables (NPIV). 126
  3.1 Exposition of the Problem ........................................... 126
  3.2 Non-Parametric Instrumental Variable (NPIV) ............... 129

4 Empirical Results; ........................................................... 140
  4.1 The data. .................................................................. 140
  4.2 The Choice of the Instrument: ..................................... 141
  4.3 Direct Estimation; ..................................................... 144
  4.4 Nonparametric Results .............................................. 149

5 Conclusion ................................................................. 155
B Appendix

4 Federal Transfers, Decentralization and the Labor Market

C Appendix

5 Consumption over the Lifecycle, joint with Jonathan Parker
4 Estimation Strategy .............................................. 230
4.1 Method of Simulated Moments (MSM) Estimation ............ 230
4.2 Remaining Calibration and Data ................................ 234
5 Estimation Results .................................................. 237
5.1 Results for the Entire Sample .................................. 238
5.2 Disaggregated evidence ......................................... 243
5.3 Robustness Checks and Extensions ............................ 245
6 Conclusion ............................................................. 249

D Appendix ............................................................. 251
 a Solving the consumer's problem ................................. 251
   a.1 Euler equation ............................................... 251
   a.2 Gauss-Hermite quadrature ................................. 251
   a.3 Consumption rules ......................................... 252
 b Data ................................................................. 253
 c Method of Simulated Moments (MSM) .......................... 255

Bibliography .......................................................... 283
List of Tables

2.1 UK-US. Prime Rate Differential .................................. 47
2.2 Canada-US. Prime Rate Differential ............................... 47
2.3 Japan-US. Prime Rate Differential ............................... 48
2.4 UK-US. Euro 3 months Differential ............................ 48
2.5 France-US. Euro 3 months Differential ......................... 49
2.6 Germany-US. Euro 3 months Differential ...................... 49
2.7 Italy-US. Euro 3 months Differential .......................... 49
2.8 Canada-US. Euro 3 months Differential ....................... 50
2.9 Japan-US. Euro 3 months Differential ......................... 50
2.10 UK-US. Prime Differential. Market Filter .................... 53
2.11 Canada-US. Prime Differential. Market Filter ............... 53
2.12 Japan-US. Prime Differential. Market Filter ................. 53
2.13 UK-US. Euro 3 months Differential. Market Filter ........... 54
2.14 France-US. Euro 3 months Differential. Market Filter ..... 54
2.15 Germany-US. Euro 3 months Differential. Market Filter ... 54
2.16 Italy-US. Euro 3 months Differential. Market Filter ...... 55
2.17 Canada-US. Euro 3 months Differential. Market Filter ..... 55
2.18 Japan-US. Euro 3 months Differential. Market Filter ..... 55
2.19 Maximum Likelihood Estimation ................................. 82
A.1 ARMA, Euro 3 months ............................................. 86
A.2 ARMA, Money Market Rates ..................................... 87
A.3 ARMA, Prime Rate ................................................. 87

3.1 EMS Regimes ......................................................... 141
3.2 Stock Market Indices .............................................. 143
3.3 Direct Estimation: Denmark ....................................... 144
3.4 Direct Estimation: Belgium ....................................... 145
3.5 Direct Estimation: France ......................................... 145
3.6 Direct Estimation: Netherlands ................................. 146
3.7 Direct Estimation: Italy ........................................... 146
3.8 Direct Estimation: Ireland ........................................ 147
3.9 Direct Estimation: Non EMS Countries ......................... 147
3.10 T-test of H0: $A_1A_2 > 0$, EMS countries .................... 148
3.11 T-test of H0: $A_1A_2 > 0$, Non-EMS countries ............... 149
3.12 Reduced Parameters, Denmark .................................... 150
3.13 Reduced Parameters, Belgium .................................... 150
3.14 Reduced Parameters, France ..................................... 151
3.15 Reduced Parameters, Netherlands .............................. 151
3.16 Reduced Parameters, Italy ........................................ 151
3.17 Reduced Parameters, Non-EMS Countries ....................... 151
3.18 Structural Parameters, Denmark .................................. 152
3.19 Structural Parameters, Belgium .................................. 152
3.20 Structural Parameters, France .................................... 153
3.21 Structural Parameters, Netherlands ............................. 153
3.22 Structural Parameters, Italy ....................................... 153
3.23 Structural Parameters, Non-EMS Countries ..................... 153

16
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Cell Sizes for Consumption Profiles</td>
<td>226</td>
</tr>
<tr>
<td>5.2 F-Tests for Flatness of Consumption Profiles</td>
<td>229</td>
</tr>
<tr>
<td>5.3 Variance of Income Shocks</td>
<td>236</td>
</tr>
<tr>
<td>5.4 Data for Retirement Consumption Rule</td>
<td>237</td>
</tr>
<tr>
<td>5.5 Structural Estimation</td>
<td>239</td>
</tr>
<tr>
<td>5.6 Benchmark Certainty Case, $\rho = 0.4897$</td>
<td>244</td>
</tr>
<tr>
<td>5.7 Estimates from the Stochastic Model, $\rho = 0.4897$</td>
<td>245</td>
</tr>
<tr>
<td>5.8 Robustness Checks, $\rho = 0.4897$</td>
<td>246</td>
</tr>
</tbody>
</table>
List of Figures

2.1.a Impulse Responses to Monetary Innovations (ratio of non-borrowed reserves) ......................................................... 93
2.1.b Impulse Responses to Monetary Innovations (ratio of non-borrowed reserves) ......................................................... 94
2.1.c Impulse Responses to Monetary Innovations (Federal Fund Rate) 95
2.1.d Impulse Responses to Monetary Innovations (Federal Fund Rate) 96
2.2 Money Market and Prime Rate ........................................... 97
2.3 Money Market and Euro 3-months Rate ............................... 98
2.4 Euro 3 months and Forecasts ........................................... 99
2.5 Delayed Overshooting ..................................................... 100
2.6 Market Belief ................................................................ 101
2.7.a Conditional Delayed Overshooting Region ....................... 102
2.7.b Conditional Delayed Overshooting Region ....................... 103
2.7.c Conditional Delayed Overshooting Region ....................... 104
2.8.a Unconditional Delayed Overshooting Region .................... 105
2.8.b Unconditional Delayed Overshooting Region .................... 106
2.8.c Unconditional Delayed Overshooting Region .................... 107
2.9 Monte Carlo Distribution of $\beta_1$ ................................. 108
2.10 Monte Carlo Simulation: $\beta_1$ vs $se(\beta_1)$ .................... 109
2.11 Money Market Differentials ........................................ 110
2.12 Innovations to Differentials ....................................... 111

3.1 SS Curve .............................................................. 160
3.2 DM-FF Log Exchange Rate and Fundamental .................. 161
3.3 Interest Rate Differential, France-Germany ................. 162
3.4.1 Compounded Excess Return Differential, EMS .......... 163
3.4.2 Compounded Excess Return Differential, Non-EMS .... 164
3.5 Exchange Rate and Fitted: Denmark ......................... 165
3.6 Exchange Rate and Fitted: Belgium ......................... 166
3.7 Exchange Rate and Fitted: France ......................... 167
3.8 Exchange Rate and Fitted: Netherlands .................. 168
3.9 Exchange Rate and Fitted: Italy ......................... 169
3.10 Exchange Rate and Fitted: non EMS ..................... 170

5.2.1 Consumption Rules ............................................... 261
5.2.2 Individual Profiles ............................................. 262
5.2.3 Expected Cash on Hand ........................................ 263
5.3.1 Household Consumption and Income over the Lifecycle 264
5.3.2 Family Size ...................................................... 265
5.3.3 Cohort and Family Adjustment ................................ 266
5.3.4 Total Liquid Assets to Income Ratio .................... 267
5.3.5 Consumption[0] and Income[+] Profiles by Education 268
5.3.6 Consumption[0] and Income[+] Profiles by Occupation 269
5.3.7 Rescaled Smoothed Consumption over the Lifecycle .... 270
5.5.1 Income, Consumption and Consumption Predicted by CEQ LCH 271
5.5.2 By Ed: Income, Consumption and Consumption Predicted by
CEQ LCH ............................................................. 272
5.5.3 By OCC: Income, Consumption and Consumption Predicted by CEQ LCH .................................................. 273
5.5.4 Simulated and Actual Consumption Profiles .................. 274
5.5.5 Simulated Consumption Profiles for different $\beta$ .......... 275
5.5.6 Savings Rate Profile ................................................. 276
5.5.7 Target Cash on Hand .................................................. 277
5.5.8 Life Cycle and Buffer Savings ...................................... 278
5.5.9 Simulated and Actual Consumption Profiles, by OCC ...... 279
5.5.10 Simulated and Actual Consumption Profiles, by Ed ........ 280
5.5.11 Simulated and Actual Consumption Profiles, Robustness Checks 281
Chapter 1

Introduction

A thesis in economics takes either the form of a dissertation or of a collection of essays. The distinction between the two resides in the delimitation of a field of investigation. According to the Webster's New International Dictionary, a dissertation consists in the extensive treatment of a particular subject. It has the advantage of consistency in the definition and analysis of its subject, an essential element of any good research. A dissertation, furthermore, tries to cover all the relevant aspects of a given problem and to provide all the ingredients necessary to its solution or understanding. The dissertation candidate, within that format, learns the essentials of research conduct, from the design of a given problem and the choice of methods, to the research agenda. This devotion of one's intellectual resources to a single objective can produce formidable results.

By contrast, a collection of essays imposes fewer constraints on its author. Within this more flexible corset a student can indulge in some experimentation. This is often reassuring as one does not need to chart from the onset a precise course of action. Eventually, one might decide to pursue the same avenue or not. At any rate, there is more to a collection of essays than alleviating graduate students' "angst." For one thing, the economic academic environment favors the production and publication of relatively short, self-contained, papers that befit the format of thesis essays. Good essays demon-
strate abilities that, too, are essential to the conduct of good research: the ability to identify and delineate precisely a valuable field of investigation, a minimalist framework that emphasizes the essential aspects of the problem under consideration and highlights the contribution to economic knowledge. The thesis candidate learns how to practice synthesis and how to structure in a limited space a potentially vast source of external knowledge. Furthermore, a collection of essays allows some degree of cross specialization that is important in a discipline like Economics. Economists are often called upon to give their opinion on real world problems, be it in informal discussions or in the course of their career. Few real economic problems have the sophistication of presenting themselves as pure as textbook problems. It is the ability to combine different degrees of analysis in an enlightening way that defines the good economist.

Ultimately, both approaches are valid research procedures, to be evaluated by the quality of the results produced. It is now to the reader to determine whether the present thesis qualifies as such.

My thesis consists in a truly heterogeneous collection of essays. This reflects my interests in International Finance and Macroeconomics, as well as “controlled historical accidents”, such as the development of joint research projects. Therefore, I will not undertake the impossible task of unifying such disparate elements.

My first essay, “Exchange Rate Dynamics and Learning,” a joint work with Aaron Tornell, seeks to understand and explain the behavior of nominal exchange rates – a paramount objective of modern International Finance. Among the many puzzling features that characterize exchange rates, the forward discount bias occupies a special place in the amount of attention it has attracted in the profession. To such extent that one might be tempted to conclude, as Ed Kane in the early 70's regarding the literature on the term structure of interest rates:

“it is generally agreed that, ceteris paribus, the fertility of a field is roughly
proportional to the quantity of manure that has been dumped upon it in the recent past. By this standard, the term structure of interest rates has become...an extraordinarily fertile field indeed."

The forward discount bias literature tests the proposition that the forward discount, as defined by the interest differential, is an unbiased predictor of the exchange rate depreciation. Although the conditions under which this proposition is strictly true are rather stringent, it proves surprisingly hard to construct sensible models that generate the type of deviations observed in the data. Put simply, not only is the forward rate a biased predictor of the future spot exchange rate, but the direction of bias contradicts the predictions of most models. Positive interest rate differentials are more likely to be followed by an appreciation of the currency than a depreciation. Equivalently, there are excess returns on the foreign exchange market. From an accounting point of view, the deviations from Uncovered Interest Parity can only come from fluctuating risk premia or expectational errors. The first avenue is by now widely dismissed as a serious candidate. The one we chose to pursue is potentially more fruitful, but also more dangerous, given the current paradigm of economic theory: how can we possibly generate large persistent expectational errors within the confines of rational models?

Our work on this topic originated with the contribution of Eichenbaum and Evans (1995). These authors look at the response of exchange rates to monetary innovations in the US. They find that, in almost all cases and for all the relevant specifications, the exchange rate appreciates gradually following a monetary contraction and then depreciates slowly over an horizon extending up to three years. This dynamic behavior is certainly consistent with the forward discount bias: over the first part of the response, the currency appreciates and there is a positive interest rate differential. Our research strategy consists in identifying the conditions under which this “delayed overshooting” could arise in equilibrium. Ultimately however, we are interested in the explanatory potential of our theory in terms of the forward discount bias. The empirical work of
Eichenbaum and Evans also identifies monetary policy as the source of this puzzling dynamics. Accordingly, we focus on the interaction between the Federal Reserve and financial markets. We develop a framework in which agents have to learn the stance of monetary policy. Every period, a new interest rate is set by the monetary authority and the market must assess the persistence of current monetary policy. Our results indicate that delayed overshooting occurs only when financial markets under-estimate the relative importance of persistent as opposed to transitory shocks to the interest rates. We then gather interest rate forecast from major industrialized countries and investigated their time-series properties of the interest rate differentials. The results conform to the predictions of the theory: market participants do indeed under-estimate the relative persistence of interest rate shocks. However, the data reveal a much more complex structure. Interest rate forecasts overpredict the transitory component, yet the perceived conditional persistence is larger. This pattern of underprediction followed by overadjustment, while unexplained in this essay, is sufficiently robust to warrant further examination. This is the object of my present research. Lastly, we are somewhat less successful at explaining away the forward discount bias. Our approach cannot generate the negative coefficients one typically finds in empirical studies.

My second essay "Non Linearities in Exchange Rates" is a follow-up on my econometrics paper, and looks at non-linearities in target zones. Target zone models of exchange rates, developed initially to analyze the behavior of exchange rates in a stochastically regulated environment, imply a non-linear relationship between the fundamental and the exchange rate. The non-linearity -whether stabilizing or not- reflects the influence of the "bands" on exchange rate expectations, and constitute therefore a crucial implication of the theory. Most direct estimation procedures construct a "fundamental", typically the exchange rate plus a fraction of the interest rate differential, and regress non-linearly the exchange rate on this variable. This immediately runs into the problem of measurement error in a non-linear specification, a delicate problem that cannot be solved by
Instrumental Variables. My original idea was to get around this using a form of inverted Maximum Likelihood estimation: if the exchange rate is in a one to one relationship with the fundamental for a given set of parameters, and the fundamental itself follows a specified stochastic process, then one could invert this functional form, extract the fundamental directly from the exchange rate observations and estimate the parameters using Maximum Likelihood. Although this method appears appealing—in particular, it only requires exchange rate observations— it never quite managed to produce sensible results, as a result of the violation of the differentiability condition of the likelihood at the extremum. This condition is necessary for the consistency and asymptotic normality of the local extremum.

Given this relatively disappointing result, I decided instead to follow an entirely different track. Seminal work by Hausman, Newey, Ichimura and Powell (1991) derives regularity conditions allowing to solve the measurement error problem in a series representation, by using an instrument and an auxiliary regression. I extend their method to a non-parametric representation, and use the compounded excess return differential as an instrument for the constructed fundamental. Applying this method allows one to recover non-linearities directly, although it does not allow one per se to test a particular target zone theory. The empirical results, performed on daily EMS data since 1979 demonstrate the existence of large non-linear effects within target zones. This approach provides a versatile estimation method that can potentially solve measurement error problems in general non-linear specifications for a variety of economic situations.

My third essay, “Federal Transfers, Decentralization and the Labor Market”, was written in the summer 1993, at the end of my first year at MIT, and echoes the European themes of the time, which I had previously analyzed in my DEA memoir “Fiscalité Fédérale et Zones Monétaires Optimales: le cas de la France”. 1991 saw the signing of the Maastricht Treaty. The European Monetary Union had not yet lost its momentum and for many economists, it became urgent to address the loss of exchange rate as a sta-
bilization instrument. These events revived an old literature dating back to the sixties, and Mundell (1961)'s seminal contribution, on Optimal Currency Areas. According to this literature, a large federal fiscal policy could offset, through the automatic stabilizer, the loss of the exchange rate. Further, empirical analysis, looking at existing monetary unions, such as the US, concluded that federal fiscal authorities play indeed a substantial role in smoothing out state specific fluctuations. This essay challenges the usefulness of the comparison between the US and Europe by emphasizing the very different structure of their labor markets. While the US has a very decentralized labor markets, Europe is characterized by national labor markets. In this context, I argue, international transfers may be captured by national pressure groups, and distort incentives to adjust to aggregate shocks. This is a standard moral hazard problem. While it sheds some pessimistic light on the plight of a post monetary unification Europe, I also show that it is nevertheless optimal to keep some federal transfers, as they allow some degree of intra-temporal risk sharing. The paper thus concludes that centralization, though limited, dominates decentralization.

Lastly, my fourth essay explores an entirely different topic: consumption theory. “Consumption over the Lifecycle” is a joint work with Jonathan Parker, a fellow graduate student at MIT. Jonathan and I took a macro course (14.453 for the initiated reader), taught by Olivier Blanchard in the fall of 1993, our first year at MIT. During the first half of the course, Olivier explored the subtle ramifications of consumption theory, with a special emphasis on prudence, precautionary saving and buffer stock theory, developed in part just recently at MIT by Christopher Carroll. Most of the buffer stock and precautionary saving literature was then addressing high frequency fluctuations in consumption, such as the violation of the linearized Euler equation or the correlation between income and consumption. Olivier suggested, by the end of the class, that one should explore the interactions between precautionary and lifecycle saving. Jonathan and I decided to do just that. Jonathan's previous experience with large micro dataset
at the University of Michigan and my natural delectation at writing Gauss code made a perfect combination. We employ a synthetic cohort technique and Consumer Expenditure Survey data to construct average age-profiles of consumption and income over the working lives of typical households across different education and occupation groups. Following recent developments in the empirical literature on lifecycle profiles, we take great care to control for possible spurious effects arising from cohort or family effects. Even after controlling for those factors, typical consumption profiles are not flat, and seem to track income at young ages. Using these profiles, we then estimate a structural model of optimal life-cycle consumption expenditures in the presence of realistic income uncertainty. Advances in computing power allowed us to perform a task absolutely out of reach just a few years ago. Thanks to the unflagging efforts by the department of economics to maintain the best computing technology, we were able to perform successive simulations quickly and accurately. Our model fits the profiles quite well. In addition to providing tight estimates of the discount rate and risk aversion, we found that consumer behavior changes strikingly over the life-cycle. Young consumers behave as "buffer-stock" agents. However, around age 43, the typical household starts accumulating liquid assets for retirement and its behavior mimics more closely that of a certainty equivalent consumer. This change in behavior is mostly driven by the life-cycle profile of expected income. This last essay provides a forceful reconciliation of the early Lifecycle models and the empirical evidence. Once realistic levels of uncertainty are introduced, and precautionary savings are incorporated, low frequency movements in consumption seem to conform to the theory. Although it does not address macroeconomic issues, it is clear that our work has important aggregate implications. It especially emphasizes the importance of demographic patterns on the aggregate saving rate. It is perhaps not simply a coincidence that the stock market is booming and national savings are rising, as the baby boomer generation starts turning 50 and, in the terminology of our paper, moves from buffer-stock behavior to a lifecycle behavior.
Chapter 2

Exchange Rate Dynamics and Learning

2.1 Introduction

This chapter presents a nominal exchange rate model with optimizing agents and learning that rationalizes two anomalies of the foreign exchange market: the forward discount puzzle and the "delayed overshooting" of exchange rates in response to monetary shocks.

The forward discount puzzle has been the subject of a large body of empirical literature (see Hodrick (1988) and Lewis (1994) for surveys). Typically a persistent pattern of exchange rate appreciation (resp. depreciation) coexists with a positive (resp. negative) differential of the domestic over the foreign interest rates. This implies that excess returns in the foreign exchange market are partially predictable. Moreover, these predictable excess returns are time-varying.

The delayed overshooting puzzle has been uncovered by Eichenbaum and Evans (1995). These authors find that unanticipated expansionary shocks to US monetary policy are followed by (a) persistent reductions in US interest rates, and (b) a gradual depreciation of the dollar, followed by a gradual appreciation several months later.
Figure 1a-d replicate the findings of Eichenbaum and Evans.1 This hump-shaped exchange rate response is a violation of the rational expectations overshooting principle (Dornbusch (1976)) whereby the exchange rate should depreciate instantly, and then appreciate gradually towards its long run equilibrium value (see Figure 5). This delayed overshooting path is potentially consistent with predictable excess returns as there exists a time interval during which a positive interest differential coexists with an appreciating currency.2

Two classes of explanations have been put forth to rationalize the forward discount puzzle: time-varying risk premia and expectational errors (see Lewis (1994) and Frankel and Rose (1994) for surveys). Under the first explanation, fluctuations in the forward rate reflect changes in the risk premium. As Fama (1984) points out, under this interpretation, the risk premium must be more volatile than predictable excess returns. In equilibrium, the risk premium will fluctuate with relative asset supplies, conditional variances, and the intertemporal elasticity of substitution of consumption. However, as with the equity premium puzzle (see Mehra and Prescott (1985)), one has to invoke unrealistically high risk aversion coefficients in order to make the risk premium fluctuations implied by the data compatible with the low volatility of the above-mentioned variables.

Using survey data on exchange rate expectations, Frankel and Froot (1989) have decomposed predictable excess returns into their risk premium and expectational error components. Their results indicate that (a) almost none of the bias can be attributed to risk premium fluctuations and (b) changes in the forward premium reflect one for one changes in expected appreciation. Thus, expectational errors are responsible for most of the bias. Expectational errors may arise either when agents have to learn about some unobservable shift in the economic environment.3 Learning about a one-time unobserv-

---

1 Appendix A describes the methodology. The results of Clarida and Gali (1994), Gilli and Roubini (1994) also point in the same direction.
2 The time dimension of this phenomenon is worth emphasizing: the nominal exchange rate peaks 10 to 36 months after the initial shock.
3 Another possibility is the presence of irrational traders. Models with irrational traders have been an-
able shock to fundamentals in the foreign exchange market has been analyzed by Lewis (1989a) and (1989b). If a change in regime occurs, agents will gradually update their beliefs about the probability that the new regime is in place, generating systematic forecast errors during the transition. These learning models explain a significant part of the exchange rate mispredictions implied by the forward discount bias. However, they do not account for the fact that predictable excess returns do not appear to die out over time between regime switches.4

Since learning generates forecast errors that die out over time, models based on learning about a one-time change in regime cannot be expected to deliver a hump-shaped impulse response of the exchange rate to monetary shocks. Another force needs to be added in order to generate delayed overshooting. In this chapter, we introduce such a force: in a stationary economy in which the equilibrium exchange rate is a function of current interest rates and expected future exchange and interest rates, hump-shaped dynamics may result from (a) the interaction of learning about the duration of monetary shocks (transitory vs. persistent) and the gradual response of interest rates to monetary

---

4Another class of models in which agents have to learn about a regime shift is the so-called “Peso problem,” whereby if an expected shift in regime does not materialize in sample, expectations will appear systematically biased to the econometrician. Kaminski (1993) shows that Peso problems can account for part of the forward discount premium in a model in which regime switches follow a Markov process. This class of models does not deliver a hump-shaped impulse response of the exchange rate to a monetary shock. Moreover, using option prices on Dollar-Deutschemarks between 1984 and 1993 to extract jump-expectations associated with shifts in regimes or bursting bubbles, Baily and Kropyviansky (1994) conclude that although there were significant jump-expectations, enough jumps were present in sample. Therefore peso problems are unlikely to have induced a major bias in exchange rate pricing.
shocks and (b) the *discrepancy* between the actual distribution of interest rate shocks in sample and its perception by market participants.

As a first step towards solving these anomalies, we analyze interest rate expectations. In order to do so, we assume that the economy is constantly hit by monetary shocks which can be temporary or permanent. Temporary monetary shocks generate transitory changes in the interest rate, while permanent monetary shocks generate persistent interest rate changes that die out gradually. Agents only observe the realization of interest rates. This specification is flexible enough and can accommodate purely transitory processes as well as persistent ones. Key parameters of this process are the speed of convergence of the persistent component and the relative variance of transitory and persistent components. Our initial task consists in estimating the forecasting rule applied by market participants. We use a unique survey data set on interest rate expectations, published by *Currency Forecasters' Digest*, with monthly observations from 1986 to 1995 for G-7 countries. We find that (a) there is no evidence of a transitory component in the *sample* distribution, while (b) market participants implicitly assume a substantial share of the shocks to be transitory. This contrast is striking: while the estimated relative variance of transitory shocks in sample is close to 0, the relative share implicitly assumed by the market is often significantly larger than one, indicating that most shocks are believed to be transitory. We emphasize that this result need not be inconsistent with rational expectations. First, the sample realization of shocks may exhibit more persistent shocks than the ergodic distribution, equivalently, there might be a Peso problem. Second, the true data generating process may shift over time, and agents may be in the continual process of "catching up". According to this interpretation, at any point in time, agents do not know the correct parameters.

Using our framework, we then develop a model of nominal exchange rate pricing that rationalizes both predictable excess returns and the delayed overshooting puzzle. We consider a typical optimizing model with a representative agent, imperfect information
and risk aversion, and derive the equilibrium exchange rate in closed form. The interest rate process is as described above. During each period agents estimate the probability distribution of future interest rates and exchange rates using the current realization of interest rates. Based on these assessments agents make their consumption and portfolio decisions in order to maximize expected utility. Lastly, the exchange rate adjusts every period in order to ensure that demand and supply of domestic bonds are equalized.

To illustrate how our model generates the delayed overshooting found by Eichenbaum and Evans consider a persistent increase in the domestic interest rate at time 0 followed by no other shock (conditional delayed overshooting). The interest rate differential jumps up at impact and returns gradually to its long run value. If agents know the exact composition of the interest rate shock, the exchange rate appreciates on impact, overshooting its long run level, and then depreciates gradually ensuring that uncovered interest parity holds. This is the standard interest rate effect characterized by Dornbusch (1976). However, if agents do not know the nature of the shock, a delayed overshooting path, as in Figure 5, may occur. This path reflects an additional learning effect. Since the actual sequence of interest rate differentials is greater than what agents expect, their assessment of the mean of future differentials goes up. This generates an increasing demand for domestic assets over time. This learning effect may dominate the interest rate effect, implying an appreciation over time in order to clear the increased demand for domestic assets. However, this appreciation cannot go on forever. As soon as market participants have gathered enough information from interest rate realizations, learning slows down, and the interest rate effect eventually dominates the learning effect. Since domestic interest rates decline toward their steady state level, the exchange rate must revert to a depreciating path at some point in time.

Uncertainty regarding the duration of the shock does not automatically generate conditional delayed overshooting. The key factors that determine whether it occurs are the speed of interest rate adjustment and the learning speed – that is, the amount of
information conveyed by current interest rates. An economy converging more rapidly to its long run equilibrium is less likely to exhibit conditional delayed overshooting. Since convergence happens very quickly, permanent monetary shocks look like transitory ones. Changes in the informativeness of the interest rate signal have more complex effects. Conditional delayed overshooting only occurs for intermediate learning speeds. With high learning speed, beliefs converge almost instantly to the true value of the persistent component of interest rates. Conversely, if learning is too slow, the market belief increases very little at the time of the shock. Afterwards, although the market belief is updated upwards, the increase in demand for domestic assets is too small to dominate the interest rate effect and there is no delayed overshooting.

Although the experiment presented above captures the basic intuition of the chapter, it is misleading along one important dimension. We described the response to a persistent shock. However, the empirical evidence presented in Eichenbaum and Evans (1995), Grilli and Roubini (1994) and Clarida and Gali (1994) does not control for the persistence of the shocks. As long as market participants use the correct distribution of shocks, there is no unconditional delayed overshooting, nor predictable excess returns. However, if persistent shocks are more frequent in sample than expected by agents, we find that unconditional delayed overshooting and predictable excess returns may occur. This case coincides with our findings using survey data on interest rates. We characterize unconditional delayed overshooting and show that empirical parameters for G-7 countries belong to the “delayed overshooting region”. Given the lack of constraints imposed on the coefficients and the simplicity of our economy, these results suggest that learning about the current state of affairs and interest rate expectations are essential components of exchange rate determination.

Finally, it is important to stress that predictable excess returns can exist without delayed overshooting. In the last section, we extend the basic model by assuming that the interest rate process includes an ARCH (autoregressive conditional heteroscedasticity)
component. This assumption captures the tendency in financial data for volatility clustering: large (small) shocks are followed by large (small) shocks. In this setup the larger the interest rate shock, the smaller the precision of the Bayesian updating. Thus, upon realization of a shock, agents’ risk premium increases, as the perceived variance of the asset is larger. Over time, as the variance and the learning speed converge back to their equilibrium values, the extra risk premium disappears. This time varying risk premium generates predictable excess returns that vary over time. However, as long as there are no expectational errors, this does not generate unconditional delayed overshooting. We then use a two step procedure to estimate our model for G-7 countries from 1974 to 1992. In a first step, we estimate the interest rate differential process by Maximum Likelihood. The results strongly support the assumption of conditional heteroscedasticity. In a second step, using a Simulated Method of Moments, we estimate the model’s remaining parameter. Having estimated our model, we use it to perform a Monte-Carlo study of deviations from Uncovered Interest Parity. We show that risk premium fluctuations related to learning account for a substantial fraction but not the entirety of the forward discount puzzle.

The structure of the chapter is as follows. In section 2 we analyze the interest rates forecasts data. In section 3 we present the basic model and specify the conditions under which delayed overshooting occurs. Section 4 assumes a conditionally heteroscedastic process for the exchange rate and shows that this implies a time-varying risk premium. We present our conclusions in section 5. Lastly, appendix A replicates the methodology behind the impulse responses of Eichenbaum and Evans (1995) for a larger time horizon, appendix B describes in more details the empirical results of section 2, while appendix C contains all algebraic derivations.
2.2 Stylized Facts About Interest Rates Differentials

We start by analyzing interest rates expectation formation. We first propose a flexible multivariate representation of the interest rates differential process. Interest rates shocks can be temporary or persistent. Temporary shocks last one period. Persistent shocks decay slowly over time. This representation allows for any pattern of positive autocorrelation of interest rate differentials, depending on the relative variance of the two primitive shocks and the speed of convergence. We estimate the parameters of the data generating process for G-7 countries against the US interest rate on monthly data. We find that (a) there is strong persistence and (b) the transitory component is negligible. This indicates that the interest rate process is well approximated by an AR process, without significant moving average part. Using survey data on interest rate forecasts at 3, 6 and 12 months horizons, we then estimate the parameters of the market filter that generate those forecasts. The market filter that best replicates the interest rates forecasts exhibits (a) a higher degree of persistence of persistent shocks and (b) a large and significant transitory component.

2.2.1 Modelling the Interest Rate Differential

We adopt a state-space representation for the interest rate differential. The interest rate differential between any two countries, $d_{it}$, is the sum of a persistent and a transitory component:

$$d_{it} = \mu + d_{it}^p + \nu_t$$

(2.1)

where $\mu$ represents a constant, $d_{it}^p$ the unobservable persistent component and $\nu_t$ is the transitory innovation. The persistent component follows an $AR(q)$ process:

$$\lambda(L) d_{it}^p = \epsilon_t$$

(2.2)
with $\lambda (L) = 1 - \sum_{i=1}^{q} \lambda_i L^i$. We assume that the transitory and persistent innovations are independent and normally distributed with mean 0 and variance $\sigma_\nu^2$ and $\sigma_\epsilon^2$ respectively. This representation is flexible enough and can accommodate an AR(q) as well as a white noise. Indeed, as we will show briefly, this representation is equivalent to a restricted ARMA process when agents do not observe individual shocks.

Following Dornbusch (1976) model we can give a monetary interpretation to this interest rate process when $q = 1$ and $\lambda = \lambda_1 < 1$. We can think of the transitory shock $\nu_t$ as a relative velocity shock, and of the persistent shock $\epsilon_t$ as caused by a permanent relative money supply shock. In the presence of sticky prices in the short run, a permanent reduction (increase) in the nominal money stock leads to an increase (reduction) in the domestic interest rate. As prices adjust over time, real money supply increases and the interest rate declines gradually until it reaches its steady state value. The speed of adjustment $\lambda$ corresponds to the root of the Dornbusch model which is less than one. A smaller $\lambda$ means faster convergence to long run equilibrium.\(^5\) This response is consistent with the empirical finding of Eichenbaum and Evans (1995): a shock to the US money supply induces a persistent change in the US interest rate in the opposite direction.

A setup with temporary and permanent monetary shocks seems appropriate to capture investors' reactions to Federal Open Market Committee (FOMC) meetings.\(^6\) One can think of each meeting as a monetary shock. Since investors only have access to the minutes of the meetings after a six weeks delay, they must conjecture whether the last decision -or lack of decision- of the FOMC will be persistent or transitory, and whether it reflects a response to inflationary pressures or concern for a weakening economy. A recent example may illustrate this point. On July 6 1995, after a two day FOMC meeting, the

\[^5\]In Dornbusch (1976), the economy converges faster the lower the semi elasticity of money demand to interest rates, and the higher the semi elasticity of output to interest rates.

\[^6\]See Batten, Blackwell, Kim, Nocera and Ozeki (1990) for a description of the operating procedures of major Central Banks.
Fed announced a cut of 25 basis points in the Federal Funds Rate, to 5.75%, ending an eighteen months long period of rising interest rates. The market was somewhat caught by surprise, as some market participants expected interest rates to remain unchanged and the bond market reacted strongly with a 1 5/8 points increase in bond prices following the announcement.\(^7\) This prompted a widespread revision of short term interest rates expectations, as traders expected further cuts, or in our terminology a persistent shock.\(^8\) In the following days and weeks, however, it appeared increasingly apparent that the economy was still moving forward. Thus, expectations of further cuts decreased, and some even feared an interest rate increase, i.e. a transitory shock.\(^9\) Furthermore, the minutes of the July 6 meeting, released on August 25, reveal that the Fed itself was surprised by the strength of the US economy. Despite being divided about the initial cut, it had planned further interest rates reduction in the weeks following the meeting.\(^10\) Thus, transitory shocks may arise when the Fed acts on inaccurate forecasts or to reflect balance of power adjustments among the Open Market Committee members. Both elements are not observed by market participants who have then to infer the motivation behind recent policy decisions.\(^11\)

Agents only observe the realization of the interest rates differential. They do not

---

\(^7\) "Just a few weeks ago, financial markets were convinced that the Fed was on the verge of cutting rates. But recent economic data have led some Fed watchers to change their mind" Wessel (1995, July 5).

\(^8\) "Now speculation turns to the future.[...]. Recent history suggests there is more to come; the Fed rarely changes direction without seeing a series of moves ahead. Every 'monetary-policy decision must be thought of as a first step along a path,' Vice Chairman Alan Greenspan said in a speech in May." Wessel (1995, July 7)

\(^9\) "After Mr. Greenspan delivered his semiannual address to the House Banking Committee, the stock and bond market concluded the Fed had no plans to cut interest rates further, traders said. Disappointed, stock and bond investors responded by selling heavily." Rebbello (1995, July 20)

\(^10\) "Federal Reserve policy makers were deeply divided about their decision to cut short-term interest rates[...] last month, the Fed disclosed. In an indication that Fed officials, at least in July, expected further reductions in rates in the months ahead, the Committee also voted to lean toward lowering rates in the weeks following the meeting" Wessel (1995, August 28).

\(^11\) Furthermore, the FOMC may grant authority to the Chairman to implement interest rates changes between meetings. Fed watchers must then be actively monitoring the New York Federal Reserve Bank’s open market operations and reserve flows (see Wessel and Raghavan (1994, March 24)).
know whether a change in the interest rates differential is caused by a transitory \((v_t)\) or persistent \((\epsilon_t)\) innovation. Similar processes have often been used in the learning literature, starting with Muth (1960) in his exploration of the link between rational and adaptive expectations. The best an agent can do is to recognize that the interest rate differential follows an \(ARMA(q,q)\). Indeed, applying \(\lambda(L)\) to (2.1), we obtain:

\[
\lambda(L) \, d_t = \epsilon_t + \lambda(L) \, v_t = \phi(L) \, u_t \tag{2.3}
\]

The right hand side is a moving average of order \(q\): \(\phi(L) = 1 + \sum_{j=1}^{q} \phi_j L^j\). The coefficients of the ARMA representation are restricted by the following system of equations, obtained by equating moments:

\[
\sigma_{\epsilon}^2 + \left(1 + \sum_{i=1}^{q} \lambda_i^2 \right) \sigma_v^2 = \left(1 + \sum_{i=1}^{q} \phi_i^2 \right) \sigma_u^2 \tag{2.4}
\]

\[
\phi_i = -\lambda_i \frac{\sigma_{\epsilon}^2}{\sigma_u^2}, \quad 1 \leq i \leq q \tag{2.5}
\]

As can be checked immediately, the moving average part disappears if and only if there is no transitory component for the interest rate differential \((\sigma_v^2 = 0)\). Since the original state-space representation has \(q + 2\) unknowns and \(ARMA(q,q)\) estimation yields \(2q + 1\) parameters, this gives \(q - 1\) overidentification restrictions.\(^{12}\) Defining the relative variance coefficient \(\eta = \sigma_v^2/\sigma_\epsilon^2\), it is easy to check that the coefficients of the moving average part \(\{\phi_i\}_{i=1}^{q}\) depend only on \(\eta\) and the \(\lambda_i\) s: \(\phi_i = \phi_i (\{\lambda_j\}_{j=1}^{q}, \eta)\). The previous decomposition suggests that one way to estimate the process (2.1)-(2.2) consists in running ARMA processes of various orders, estimating \(\{\phi_i, \lambda_i\}_{i=1}^{q}\), and inverting (2.4)-(2.5), imposing the \(q - 1\) overidentifying restrictions. There are three problems with this approach. First, this is not efficient if the model is well specified. Second, it is

\(^{12}\)When \(q = 1\), the model is exactly identified. However, \(\lambda\) and \(\phi\) must still be of opposite signs.
unclear how one should choose the order of the autoregressive part. This problem is particularly relevant when there is no transitory component so that the process follows an AR(g). In practice, we picked the ARMA process minimizing Akaike's criterion. The results are reported in appendix B. In most cases, this identification procedure failed to detect a moving average component.\footnote{Two exceptions are UK-US and Germany-US for the Euro-3 months rate (Table A.1). Both cases violate the restriction imposed by our assumptions. Moving average components are detected for France-US and Italy-US using the 1 month money market rates.} Lastly, the state-space representation is more appropriate to make exact finite sample forecasts. We will need such forecast to estimate the parameters of the distribution implicitly assumed by market participants.

The results from the identification procedure are reported in appendix B, Table A.1, A.2, and A.3. The results are consistent with our state-space representation and indicate that transitory components are small or absent.

2.2.2 Survey Data

While survey data on monthly money market rates are available for the US, and have been used in previous studies, we were unable to find similar survey forecasts for foreign countries.\footnote{Froot (1989) uses quarterly data on the three months T-bill from 1969 to 1986 from the Nagan Bond and Money Market Letter. This dataset has also been used in Friedman (1980).} Instead, we obtained interest rate forecasts from the Currency Forecasters' Digest, now called Financial Times Currency Forecaster. Contributors include multinational companies as well as forecasting services from major investment banks, i.e. the most active player on the fixed income and foreign exchange markets.\footnote{The Forecasting services that contribute to the Currency Forecaster's Digest are: Predex, Merrill Lynch, Mellon Bank, Harris Trust, Bank of America, Morgan Grenfell, Chase Manhattan, Royal Bank of Canada, Midland Montagu, Generale de Banque, MMS International, Chemical Bank, Union Bank of Switzerland, Multinational Computer Models, Goldman Sachs International, Business International, M. Murenbeeld, and Westpac Bank. The multinational companies that contribute are: General Electric, Du Pont, WR Grace, Allied Signal, Monsanto, Ingersoll-Rand, General-Motors, Data General, Eli Lilly, Aetna, American Express, Johnson & Johnson, Sterling Drug, Firestone, 3M, Union Carbide, Texaco, United Brands, SmithKline Beckman, American National Can, RJ Reynolds, Colgate-Palmolive, Warner-Lambert, Schering-Plough, Quaker Oats, Beatrice Foods, Hercules, Baxter Travenol, and Interpublic Group.} This monthly
publication collects interest rates and their forecasts 3, 6 and 12 months hence for the prime rate, three months and one year Eurodollar-rates and ten year government bonds. It then reports a "market average" weighting individual respondents according to their relative importance. The countries covered are Australia, Canada, France, Germany, Italy, Japan, Switzerland, UK and US. We will restrict ourselves to the G-7 countries Canada, France, Germany, Italy, Japan, UK and the US. The period for which interest rate forecasts are provided is 1986-1995. This dataset is unique in its coverage and consistency. We have not found any other source of interest rate forecasts prior to 1986 covering all G-7 countries.

We emphasize that one should be cautious when using survey data. First, there is probably no such thing as a "market expectation". Since Currency Forecaster's Digest aggregates individual forecasts according to some proprietary and non-disclosed rule, one should exert extra caution. We do not believe, however, that individual respondents have an incentive to misreport their forecasts. Each participant, in exchange for her own contribution which remains secret, gets back the market average, incorporating in an unspecified way this same contribution. This may induce more truthful revelation and mitigate incentives either to herd and discard one's private information or to report "extreme" forecasts. As an alternative to survey data, we could have instead used forward interest rates as implied by the term structure of interest rates. Under the rational expectation theory of the term structure of interest rates, the forward rates is equal to the expected future short-term interest rate. Froot (1989), using survey data on US interest rates finds that the expectation hypothesis fails at short maturities (less than 12 months), an indication that forward rates incorporate time-varying risk premium. Therefore, as a first attempt, it seems more appropriate to use directly survey data.

Prime rates are short term lending rates to preferred customers and are available from September 1987 to September 1995. They are at the bottom of the lending structure and

---

16 Currency Forecaster's Digest does not disclose its aggregation rule nor individual forecasts.
reflect monetary policy as well as market structure in the banking industry. In particular, the prime is not a market rate. Thus, one would suspect that shocks to the prime rate might originate from other sources than monetary innovations. Similarly, banks may react with some delay to innovations in money market rates or T-bill rates. Since the difference between the money market rate and the prime determines the banking margin, banks might try to keep the prime constant for a while after a cut in the discount rate, so as to increase their margin. Market structure determines the extent of the discrepancy. For all these reasons, the prime is likely to be a noisy indicator of monetary policy. In addition, the prime is changed only infrequently. Figure 2a–g show the prime rate together with the money market rate. It is clear that the prime is close enough to the market rate for the US, the UK, Canada and Japan. For France, Germany and Italy however, the prime exhibits substantial inertia and appears disconnected from the money market rates, at least for part of the sample, so that our modelling assumptions seem inappropriate. We run the estimation using the prime rate only for the US, the UK, Canada and Japan.

Instead of using prime rates, we can use 3-months rates on Euro deposits. These rates, and their forecasts at 3, 6 and 12 months are available from Currency Forecaster’s Digest from August 1986 to October 1995. Euro rates are not likely to reflect immediately the stance of domestic monetary policy. First they have a longer maturity than money market rates. Second, these are offshore rates that can be shielded from country risk and may also reflect less accurately domestic monetary conditions. Figure 3a–g reports the euro rates (annualized) against the money market rates. It appears that in most cases the correlation is quite good. Therefore, we keep all G-7 countries in our sample.

17 The money market rate is line 160b from the International Financial Statistics tape.
18 All rates except on the Euro Pound are London interbank offer rates. The Euro Pound rate is a Paris interbank offer rate.
2.2.3 Maximum Likelihood Estimation of the Interest Rate Process

We develop in this section the estimation procedure for our state-space representation, directly imposing the restrictions (2.4)-(2.5). This will prove useful when estimating the market parameters. We rewrite (2.1)-(2.2) as:

\[
\begin{align*}
\dot{d}_t &= \mu + H'\xi_t + \nu_t \\
\xi_t &= F\xi_{t-1} + \epsilon_t
\end{align*}
\]  

(2.6) (2.7)

where \(\xi_t = (d_{t}^0, \ldots, d_{t-q+1}^0)'\), \(H' = (1, 0, \ldots, 0)'\) is a \(q \times 1\) vector. \(\xi_t\) is the state vector for the process, (2.6) the measurement equation and (2.7) the space equation. Defining the informations set \(I_t = \{d_{t-j}, j \geq 0\}\) and \(\hat{\xi}_{t+1|t} = E[\xi_{t+1}|I_t]\), we can derive the filter and the smoother.\(^{19}\) Under the normality assumption, and assuming additionally that \(\hat{\xi}_{1|0}\) is normally distributed, \(\xi_{t+1}\) is normally distributed conditionally on \(I_t\), with mean \(\hat{\xi}_{t+1|t}\) and variance \(\hat{\sigma}_{t+1|t}\). We can then write the conditional likelihood of \(i_{t+1}\) as:

\[
\log f_{di_{t+1}|I_t}(i_{t+1}|I_t) \propto \log \left| H'\hat{P}_{t+1|t}H + \sigma^2_v \right| + \left( \frac{\left( d_{t+1} - \mu - H'\hat{\xi}_{t+1|t} \right)^2}{H'\hat{P}_{t+1|t}H + \sigma^2_v} \right)
\]  

(2.8)

We maximize the sample log likelihood \(\sum_{t=0}^{T-1} \log f_{di_{t+1}|I_t}(i_{t+1}|I_t)\) with respect to the vector of parameters \(\theta = (\{\lambda_i\}_{i=1}^{p}, \eta, \sigma^2_v, \mu)'\).\(^{20}\) To initiate the estimation procedure, we need an estimate of the space variable \(\hat{\sigma}_{0}^{2}\) and its conditional mean square error. Maximum likelihood estimation over the vector \(\theta\) is then performed. Once an estimate \(\hat{\theta}^0\) is found, we run the smoother in order to revise the initial state vector. That is, the smoother gives us the initial value of the persistent component, conditional on the

\(^{19}\)See appendix B and Hamilton (1994, chapter 13).

\(^{20}\)Eurorates estimation was modified to take into account a maturity larger than the sampling frequency.
entire sample information and the filter parameters, \( \hat{i}_0^{p1} = E \left[ \hat{i}_0^p | I_T, \hat{\theta}^0 \right] \), and its mean square error. In general, this revised estimate does not correspond to the initial one. We can then iterate the maximum likelihood estimation with this new initial state variable until convergence to \( \hat{\theta}^1 \). Iterating this procedure will give ultimately a parameter vector consistent with the initial state vector.\(^{21}\)

We report in this section maximum likelihood estimation for the prime rate for the UK, Canada, Japan versus the US and the Euro 3-months rates for all G-7 countries against the US.

Results for the prime rate, reported in Tables 2.1, 2.2 and 2.3, indicate that persistent shocks disappear extremely slowly. The long run autocorrelation is always above 0.94. For UK-US and Japan-US, the long run autocorrelation is not statistically different from 1, indicating a possible unit root in the persistent component of the interest rate differential. Although the standard errors reported are often incorrect, small sample forecasts based on the point estimates are still correct. These extremely high values for the long run autocorrelation may reflect the relative inertia of the prime rate.\(^{22}\)

The implied annual serial correlation ranges from 0.54 to 0.94. The relative variance parameter is never statistically different from 0.\(^{23}\) This confirms the ARMA results and suggests that transitory components are not present. Since these results might be the consequence of the relative inertia of the prime rates (see Figure 2a – g), we present next the results for the Euro 3-months interest rates estimated over the entire floating period in Tables 2.4, 2.5, 2.6, 2.7, 2.8, and 2.9.\(^{24}\)

\(^{21}\)The asymptotic properties are the same whether we iterate or not.

\(^{22}\)Our framework assumes a shock every period. If the interest rate differential is changed only infrequently, we are likely to spuriously estimate higher persistence.

\(^{23}\)The estimation procedure imposes strictly positive standard errors. When the transitory component is strictly 0, the filter may have difficulties working correctly. When this happens, we directly estimate an AR(\(q\)) process on the data. A value of 0 is then reported for \( \eta \).

\(^{24}\)The euro 3-months reported in the IFS tape (lines 60ldd, and 60ea) is identical to the Currency Forecaster's Digest over the 1986-1995 period. The prime rate, on the other hand, is defined differently in the IFS tape and in the survey dataset.
Table 2.1: UK-US. Prime Rate Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th></th>
<th>AR2</th>
<th></th>
<th>AR3</th>
<th></th>
<th>AR4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9957</td>
<td>(0.1026)</td>
<td></td>
<td>1.3397</td>
<td>(0.0050)</td>
<td></td>
<td>1.5217</td>
<td>(0.1037)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td></td>
<td>-0.3505</td>
<td>(0.0034)</td>
<td></td>
<td>-0.7466</td>
<td>(0.1054)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2237</td>
<td>(0.0737)</td>
<td></td>
<td>0.0628</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0617</td>
<td>(0.1043)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0000</td>
<td>(0.0211)</td>
<td></td>
<td>0.0022</td>
<td>(0.1916)</td>
<td></td>
<td>0.0056</td>
<td>(0.1092)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9957</td>
<td>(0.1026)</td>
<td></td>
<td>0.9891</td>
<td>(0.0039)</td>
<td></td>
<td>0.9987</td>
<td>(0.1796)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.0773</td>
<td></td>
<td>-0.0689</td>
<td></td>
<td>-0.0608</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. 98 monthly observations.

Table 2.2: Canada-US. Prime Rate Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th></th>
<th>AR2</th>
<th></th>
<th>AR3</th>
<th></th>
<th>AR4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9635</td>
<td>(0.0451)</td>
<td></td>
<td>1.0248</td>
<td>(0.0379)</td>
<td></td>
<td>1.0439</td>
<td>(0.0343)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td></td>
<td>-0.0788</td>
<td>(0.0347)</td>
<td></td>
<td>-0.2118</td>
<td>(0.0484)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1225</td>
<td>(0.0602)</td>
<td></td>
<td>0.3876</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.1336</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0000</td>
<td>(0.2382)</td>
<td></td>
<td>0.0000</td>
<td>(0.1077)</td>
<td></td>
<td>0.0000</td>
<td>(0.5501)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9635</td>
<td>(0.0451)</td>
<td></td>
<td>0.9460</td>
<td>(0.0277)</td>
<td></td>
<td>0.9546</td>
<td>(0.0331)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.0759</td>
<td></td>
<td>-0.0758</td>
<td></td>
<td>-0.0748</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. 98 monthly observations.
Table 2.3: Japan-US. Prime Rate Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9956</td>
<td>1.6867</td>
<td>1.5517</td>
<td>1.2926</td>
</tr>
<tr>
<td></td>
<td>(0.0126)</td>
<td>(0.1031)</td>
<td>(0.0030)</td>
<td>(6.8990)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.6940</td>
<td>-0.4308</td>
<td>-0.1316</td>
<td>-0.0818</td>
</tr>
<tr>
<td></td>
<td>(0.1031)</td>
<td>(0.0050)</td>
<td>(0.0060)</td>
<td>(6.8391)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td>-0.1316</td>
<td></td>
<td>-0.1030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0060)</td>
<td></td>
<td>(6.7745)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td>-0.0818</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0065)</td>
<td>(6.6674)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0000</td>
<td>0.0115</td>
<td>0.0110</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td>(0.1576)</td>
<td>(0.0599)</td>
<td>(0.0189)</td>
<td>(0.1520)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9956</td>
<td>0.9927</td>
<td>0.9893</td>
<td>0.9914</td>
</tr>
<tr>
<td></td>
<td>(0.0126)</td>
<td>(0.1458)</td>
<td>(0.0034)</td>
<td>(0.2699)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.0143</td>
<td>0.0044</td>
<td>0.0056</td>
<td>0.0071</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. 98 monthly observations.

Table 2.4: UK-US. Euro 3 months Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9314</td>
<td>1.1029</td>
<td>1.1404</td>
<td>1.1578</td>
</tr>
<tr>
<td></td>
<td>(0.0395)</td>
<td>(0.1006)</td>
<td>(0.0263)</td>
<td>(0.0282)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.1819</td>
<td>-0.4372</td>
<td>-0.2360</td>
<td>-0.0746</td>
</tr>
<tr>
<td></td>
<td>(0.1017)</td>
<td>(0.0262)</td>
<td>(0.0198)</td>
<td>(0.0197)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td>0.2360</td>
<td></td>
<td>0.3197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0198)</td>
<td></td>
<td>(0.0312)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td>-0.0746</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0197)</td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.4379)</td>
<td>(0.1483)</td>
<td>(0.1525)</td>
<td>(0.1996)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9314</td>
<td>0.9210</td>
<td>0.9392</td>
<td>0.9346</td>
</tr>
<tr>
<td></td>
<td>(0.0395)</td>
<td>(0.0401)</td>
<td>(0.0203)</td>
<td>(0.0232)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.3994</td>
<td>-0.3905</td>
<td>-0.3776</td>
<td>-0.3768</td>
</tr>
</tbody>
</table>

### Table 2.5: France-US. Euro 3 months Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.8616</td>
<td>(0.0743)</td>
<td>0.8602</td>
<td>(0.2269)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.0003</td>
<td>(195.1985)</td>
<td>-0.0332</td>
<td>(0.0922)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>0.0349</td>
<td>(0.0532)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td></td>
<td>0.0362</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0006</td>
<td>(0.3834)</td>
<td>0.0002</td>
<td>(0.1309)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.8616</td>
<td>(0.0743)</td>
<td>0.8600</td>
<td>(0.3176)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.6413</td>
<td></td>
<td>-0.6409</td>
<td></td>
</tr>
</tbody>
</table>


### Table 2.6: Germany-US. Euro 3 months Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9654</td>
<td>(0.0271)</td>
<td>1.2350</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.2790</td>
<td>(0.1307)</td>
<td>-0.6337</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>0.2876</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.0000</td>
<td>(0.7172)</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9654</td>
<td>(0.0271)</td>
<td>0.9561</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.3722</td>
<td></td>
<td>-0.3555</td>
</tr>
</tbody>
</table>


### Table 2.7: Italy-US. Euro 3 months Differential

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9595</td>
<td>(0.1043)</td>
<td>1.6604</td>
<td>(0.0061)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.6879</td>
<td>(0.0053)</td>
<td>0.3131</td>
<td>(0.0161)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td>-0.3955</td>
<td>(0.0114)</td>
<td>0.0551</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td>-0.2820</td>
<td>(0.0252)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.2785</td>
<td>(0.3920)</td>
<td>2.6235</td>
<td>(1.3933)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9595</td>
<td>(0.1043)</td>
<td>0.9725</td>
<td>(0.0068)</td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.3076</td>
<td>-0.2932</td>
<td>-0.2935</td>
<td>-0.2835</td>
</tr>
</tbody>
</table>

Table 2.8: CANADA-US. EURO 3 MONTHS DIFFERENTIAL

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th></th>
<th>AR2</th>
<th></th>
<th>AR3</th>
<th></th>
<th>AR4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>λ1</td>
<td>0.9019</td>
<td></td>
<td>1.1407</td>
<td></td>
<td>0.9502</td>
<td></td>
<td>1.6924</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0993)</td>
<td></td>
<td>(0.0483)</td>
<td></td>
<td>(0.1011)</td>
<td></td>
<td>(0.0119)</td>
<td></td>
</tr>
<tr>
<td>λ2</td>
<td>-0.2351</td>
<td></td>
<td>0.0387</td>
<td></td>
<td>-0.1051</td>
<td></td>
<td>-0.9306</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0355)</td>
<td></td>
<td>(0.0579)</td>
<td></td>
<td>(0.0388)</td>
<td></td>
<td>(0.0120)</td>
<td></td>
</tr>
<tr>
<td>λ3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0251</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0153)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>η</td>
<td>0.0001</td>
<td></td>
<td>0.1785</td>
<td></td>
<td>0.0002</td>
<td></td>
<td>1.2431</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5474)</td>
<td></td>
<td>(0.8295)</td>
<td></td>
<td>(0.6863)</td>
<td></td>
<td>(2.4798)</td>
<td></td>
</tr>
<tr>
<td>λ(1)</td>
<td>0.9019</td>
<td></td>
<td>0.9056</td>
<td></td>
<td>0.8837</td>
<td></td>
<td>0.9767</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0993)</td>
<td></td>
<td>(0.0378)</td>
<td></td>
<td>(0.0577)</td>
<td></td>
<td>(0.0360)</td>
<td></td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.1443</td>
<td></td>
<td>-0.1413</td>
<td></td>
<td>-0.1395</td>
<td></td>
<td>-0.1196</td>
<td></td>
</tr>
</tbody>
</table>


Table 2.9: JAPAN-US. EURO 3 MONTHS DIFFERENTIAL

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th></th>
<th>AR2</th>
<th></th>
<th>AR3</th>
<th></th>
<th>AR4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>λ1</td>
<td>0.9347</td>
<td></td>
<td>1.3062</td>
<td></td>
<td>1.1159</td>
<td></td>
<td>1.4934</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0167)</td>
<td></td>
<td>(0.0191)</td>
<td></td>
<td>(0.0707)</td>
<td></td>
<td>(0.7733)</td>
<td></td>
</tr>
<tr>
<td>λ2</td>
<td>-0.3711</td>
<td></td>
<td>-0.1063</td>
<td></td>
<td>-0.0825</td>
<td></td>
<td>-0.6696</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td></td>
<td>(0.0604)</td>
<td></td>
<td>(0.0707)</td>
<td></td>
<td>(1.3582)</td>
<td></td>
</tr>
<tr>
<td>λ3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1170</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.8772)</td>
<td></td>
</tr>
<tr>
<td>η</td>
<td>0.0000</td>
<td></td>
<td>0.1438</td>
<td></td>
<td>0.0003</td>
<td></td>
<td>0.3024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9909)</td>
<td></td>
<td>(0.9975)</td>
<td></td>
<td>(5.5217)</td>
<td></td>
<td>(0.6574)</td>
<td></td>
</tr>
<tr>
<td>λ(1)</td>
<td>0.9347</td>
<td></td>
<td>0.9350</td>
<td></td>
<td>0.9271</td>
<td></td>
<td>0.9496</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0167)</td>
<td></td>
<td>(0.0153)</td>
<td></td>
<td>(0.1225)</td>
<td></td>
<td>(0.0234)</td>
<td></td>
</tr>
<tr>
<td>Log Lik.</td>
<td>-0.3845</td>
<td></td>
<td>-0.3723</td>
<td></td>
<td>-0.3725</td>
<td></td>
<td>-0.3722</td>
<td></td>
</tr>
</tbody>
</table>

These tables confirm earlier results: there is no significant transitory component in the interest rate differential. The long run persistence is smaller, as we guessed, ranging from 0.86 (France-US) to 0.97 (Italy-US). The short run autocorrelation is frequently higher than one, indicating further deviations from equilibrium 3 months after the initial shock. The relative variance is never significantly different from zero. In summary, interest rate differentials exhibit strong persistence and no significant transitory component. This result appears robust to the order of the autoregressive process.

### 2.2.4 Estimating the Market filter

We now turn to the estimation of the market filter. We assume that agents use linear forecasting formulas, as summarized by the Kalman filter equations. However, it is unlikely that agents know the exact parameters driving the interest rate differential process. According to one interpretation, these parameters may be time-varying and agents may be in the permanent process of revising their estimates. More generically, we adopt an agnostic view and will estimate the parameters of the filter implicitly used, which we denote the “market filter”. Denote \( \hat{\theta} = (\{\hat{\lambda}_i\}_{i=1}^p, \hat{\eta}, \hat{\sigma}_2^2) \) the parameters of the market filter. For a given market filter we can generate the associated forecasts at any horizon. Suppose that the current estimate for the state variable at time \( t \) is \( \hat{\xi}_{\xi\xi\xi\xi} \).

According to (2.6)-(2.7), the market forecast for the interest rate differential \( \tau \) periods from now is:

\[
dt^\tau_t (\hat{\theta}) = E[i_{t+\tau}|I_t] = \mu + H'F^\tau\hat{\xi}_{\tau\xi|\xi} = \mu + H'F^\tau\hat{\xi}_{\tau|\xi}
\]

(2.9)

The forecast constructed in such a way uses only information up to time \( t \). Under the assumption that the reported forecast is measured with error, i.e.:

---

25Note that the market filter refers to the parameters of the filter used implicitly by market participants, not to the actual data-generating parameters in sample.
\[ d\hat{\alpha}_t = d\hat{\alpha}_t^r + \nu_t \]

and that the measurement error is uncorrelated with the true forecast, we can estimate \( \hat{\theta} \) by minimizing:

\[
S(\hat{\theta}) = \sum_{r=1}^{T} \sum_{i=1}^{T} (d\hat{\alpha}_t^r - d\hat{\alpha}_t(\hat{\theta}))^2
\]

where the summation is over both observations and forecasts horizons.26 We report results for the prime rate first in Tables 2.10, 2.11 and 2.12, then for the euro 3-months rate in Tables 2.13, 2.14, 2.15, 2.16, 2.17 and 2.18.

The results indicate that the market overestimates the speed of convergence of persistent shocks. While the long run autocorrelation is close to 0.9 in sample, the market estimates much higher long run autocorrelation, sometimes significantly in excess of 1.27 This extra persistence is compensated by a large -and often significant- estimate for the relative variance. Market participants always attributes a sizeable share of the shocks to transitory components, even though we have failed to find such components in sample. Without the excess transitoryness, interest rate shocks would be expected to die out at a much slower rate than observed in the data. However, the presence of the transitory component introduces a dampening effect: a share of the shock disappears extremely rapidly, while the rest slowly decays. We report on Figure 4 the actual forecasts and the fitted values for the dollar-pound interest rate differential. This figure indicates that our estimated market filter accurately reproduces the dynamics of the market forecasts.

The results are also strongly supportive of the model. For almost all countries and

\[ ^{26}\text{Eurorates estimation reports robust standard errors since the horizon is larger than the sampling frequency.} \]
\[ ^{27}\text{While this indicates that the process might not be perceived as second order stationary, finite sample forecasts are still well defined.} \]
### Table 2.10: UK-US. Prime Differential. Market Filter

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9798</td>
<td>(0.0027)</td>
<td>1.7227</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.7301</td>
<td>(0.0288)</td>
<td>0.6831</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.5984</td>
<td>(0.0919)</td>
<td>2.6759</td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>0.2222</td>
<td>(0.5703)</td>
<td>9.8255</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9798</td>
<td>(0.0027)</td>
<td>0.9926</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. Forecasts at 3, 6 and 12 months. 294 monthly observations.

### Table 2.11: Canada-US. Prime Differential. Market Filter

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9938</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.8876</td>
<td>(0.0646)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.6873</td>
<td>(0.0284)</td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>2.6622</td>
<td>(1.4979)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9938</td>
<td>(0.0022)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. Forecasts at 3, 6 and 12 months. 294 monthly observations.

### Table 2.12: Japan-US. Prime Differential. Market Filter

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9901</td>
<td>(0.0027)</td>
<td>1.6523</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.6564</td>
<td>(0.1710)</td>
<td>0.7930</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>-0.5554</td>
<td>(0.0128)</td>
<td>2.5929</td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>0.6768</td>
<td>(0.1092)</td>
<td>12.2833</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9901</td>
<td>(0.0027)</td>
<td>0.9959</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Sample period: September 1987 to October 1995. Forecasts at 3, 6 and 12 months. 294 monthly observations.
Table 2.13: UK-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9774 (0.0024)</td>
<td>1.7154 (0.0369)</td>
<td>2.0440 (0.0586)</td>
<td>1.2901 (0.0878)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.7229 (0.0361)</td>
<td>-1.2918 (0.1119)</td>
<td>-0.8923 (0.1444)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td>0.2427 (0.0547)</td>
<td>1.2420 (0.1042)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td>-0.6570 (0.0503)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>0.5970 (0.2077)</td>
<td>16.9190 (1.0496)</td>
<td>37.5876 (14.2598)</td>
<td>2.4625 (0.6937)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9774 (0.0024)</td>
<td>0.9925 (0.0010)</td>
<td>0.9949 (0.0006)</td>
<td>0.9829 (0.0016)</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.

Table 2.14: FRANCE-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>1.0108 (0.0025)</td>
<td>0.0958 (0.2238)</td>
<td>-0.8612 (0.0025)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td>0.9248 (0.2274)</td>
<td>0.8454 (0.3426)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>1.0122 (0.2519)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>211.9981 (26.9606)</td>
<td>57.6443 (13.8397)</td>
<td>8.3256 (0.9336)</td>
<td></td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>1.0108 (0.0025)</td>
<td>1.0207 (0.035)</td>
<td>0.9964 (0.044)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.

Table 2.15: GERMANY-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9794 (0.0017)</td>
<td>1.7322 (0.0276)</td>
<td>0.8082 (0.0017)</td>
<td>0.3564 (0.0942)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.7393 (0.0272)</td>
<td>0.8287 (0.1236)</td>
<td>0.8103 (0.1252)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td>-0.6508 (0.0453)</td>
<td>0.4081 (0.1290)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td>-0.5970 (0.0818)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\eta}$</td>
<td>0.0813 (0.1236)</td>
<td>11.0333 (3.8241)</td>
<td>2.2680 (1.6464)</td>
<td>0.6299 (0.2793)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9794 (0.0017)</td>
<td>0.9929 (0.0038)</td>
<td>0.9861 (0.0034)</td>
<td>0.9782 (0.0020)</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.
Table 2.16: ITALY-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>1.0221</td>
<td>(0.0039)</td>
<td>0.1443</td>
<td>(1.0651)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.8971</td>
<td>(1.0759)</td>
<td>-1.8203</td>
<td>(0.0578)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>1.1144</td>
<td>(0.0262)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>28.2254</td>
<td>(3.3153)</td>
<td>79.0868</td>
<td>(8.3929)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>1.0221</td>
<td>(0.0039)</td>
<td>1.0414</td>
<td>(0.0139)</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.

Table 2.17: CANADA-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9721</td>
<td>(0.0049)</td>
<td>0.2758</td>
<td>(0.1578)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.6756</td>
<td>(0.1531)</td>
<td>0.6285</td>
<td>(0.9276)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>-0.5774</td>
<td>(0.045)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td>2.9905</td>
<td>(0.9276)</td>
<td>0.7360</td>
<td>(0.3445)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9721</td>
<td>(0.0049)</td>
<td>0.9514</td>
<td>(0.0001)</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.

Table 2.18: JAPAN-US. EURO 3 MONTHS DIFFERENTIAL. MARKET FILTER

<table>
<thead>
<tr>
<th></th>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.9815</td>
<td>(0.0023)</td>
<td>1.5692</td>
<td>(0.0350)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.5787</td>
<td>(0.0342)</td>
<td>0.7970</td>
<td>(0.1800)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>-0.5449</td>
<td>(0.0078)</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.2903</td>
<td>(0.1800)</td>
<td>3.9028</td>
<td>(2.4011)</td>
</tr>
<tr>
<td>$\lambda(1)$</td>
<td>0.9819</td>
<td>(0.0023)</td>
<td>0.9905</td>
<td>(0.0012)</td>
</tr>
</tbody>
</table>

Note: Robust standard errors in parentheses. Sample period: August 1986 to October 1995. Forecasts at 3, 6 and 12 months. 333 monthly observations.
specification, the noise to signal ratio is large and significant. This implies a substantial share allocated to the transitory component.

The overall results are striking: (a) there is no significant or systematic transitory component of interest rates. This is true for all countries in the sample, and for different time periods, while (b) the parameters of the filter that best replicates interest rate expectations exhibit a substantial transitory components. This is true also across countries, and for different interest rates measures.

2.3 A Representative Agent Model

In this section we build on our empirical findings and develop a model of exchange rate determination that rationalizes both predictable excess returns and delayed overshooting. We present a model of a world economy with two countries, one consumption good, a domestic bond, and a foreign bond. The economy is populated by atomistic agents who live two periods and derive utility from consumption. Total population is constant across time. Each agent of generation $t$ is born with an endowment $W_t$ of the consumption good. It uses this endowment in consumption, purchasing $x_t$ units of the domestic bond which has a price $f_t$ in terms of the consumption good, and $W_t - c_t - f_t x_t$ units of the foreign bond which has a price of one in terms of the consumption good. The domestic bond is risky. At time $t+1$ it pays $i_t$ units of the consumption good, and is sold at a price $f_{t+1}$ (to the next generation). At time $t$, $i_t$ is known, while $f_{t+1}$ is not. The foreign bond is riskless. At time $t+1$ its price is 1 and it pays $i^*$ units of the consumption good. When old, at time $t+1$, the representative agent consumes all its wealth. It follows that

$$c_{t+1} = [W_t - c_t - f_t x_t] [1 + i^*] + [i_t + f_{t+1}] x_t$$  \hspace{1cm} (2.10)

In this setup, the exchange rate is the price of the foreign bond in terms of the domestic bond. Since the price of the foreign bond in terms of the consumption good is one, $1/f_t$
is the exchange rate.

A peculiarity of budget constraint (2.10) is that \( f_t \) and \( f_{t+1} \) enter separately. This specification combined with a constant absolute risk aversion utility function will allow us to derive the equilibrium exchange rate in closed form. This will make it possible to determine whether delayed overshooting occurs.

Following the decomposition adopted in the previous section we assume that the domestic interest rate has three components: its steady state level \( i^* \), a persistent component \( i^p_t \), and a transitory shock \( \nu_t \):

\[ i_t = i^* + i^p_t + \nu_t \tag{2.11} \]

The persistent component has two elements: the effects on \( i_t \) of past persistent shocks which die out at rate \( \lambda \), and the effects of contemporaneous persistent shocks \( \epsilon_t \):

\[ i^p_t = \lambda i^p_{t-1} + \epsilon_t \tag{2.12} \]

We assume that the shocks \( \nu_t \) and \( \epsilon_t \) are independent normal variables with 0 mean and respective variance \( \sigma^2_{\nu} \) and \( \sigma^2_{\epsilon} \). Agents in the market only observe the realization of interest rates (\( i_t \) and \( i^* \)). They do not know whether a change in the domestic interest rate is caused by a transitory (\( \nu_t \)) or a persistent shock (\( \epsilon_t \)). All agents receive the same information and draw the same conclusions from it.\(^{28}\)

We postulate that the objective function of the representative young investor at time \( t \) is CARA:

\[ U(c_t, c_{t+1}) = -\exp(-\gamma c_t) - \beta E \left[ -\exp(-\gamma c_{t+1}) \mid I_t \right] \tag{2.13} \]

where \( \gamma \) represents the coefficient of absolute risk aversion. The information set at time

\(^{28}\)This way of modelling the learning problem has a long standing in the economics literature. It was used by Mussa (1975) to show that it is rational to form expectations of future inflation using an adaptive mechanism, and by Lucas (1973) in his analysis of the Phillips curve.
$t$, $I_t$, includes all past and current interest rates and prices. That is, $I_t = \{I_{t-1}, i_t, f_t\}$. This information set is common to all agents. This is an appropriate assumption since macroeconomic variables are the relevant ones in the foreign exchange market. It is unlikely that market participants would possess asymmetric information regarding these variables for extended periods of time.

We choose a constant absolute risk aversion specification in order to solve independently the portfolio decision and the savings decision. This will allow us to obtain a linear closed form solution for the equilibrium price of the domestic bond. The specification is standard in the finance literature, and all its inconveniences are well known.

The timing is as follows. At time $t$, after the realization of the domestic interest rate $i_t$ is revealed, the representative young investor chooses $x_t$ and $c_t$ in order to maximize her expected utility (2.13). In so doing she takes as given the current and expected future price of the domestic bond $f_t$ and $E(f_{t+1}|I_t)$. In forming her expectation about $f_{t+1}$, each young investor uses her estimate of the probability distribution of $i_{t+1}$.\footnote{As we will show, in this representative agent model, there is nothing to be learned from prices. In more general finance models with information heterogeneity, each investor would extract information from current prices. Prices might then be fully revealing if markets are complete. To prevent this, finance models usually assume that supply varies stochastically (Grossman (1976), Hellwig (1980)). This assumption is not necessary in our set-up as prices are uninformative.} Lastly, at time $t + 1$ the old representative investor sells her holdings of the domestic bond to the “$t + 1$ young”, consumes all her wealth and transmits her information to her offspring.

### 2.3.1 The Learning Problem

Using survey data we found in section 2 that shocks to the interest rate differential are considered by market participants to be more transitory than what they actually are in sample. In all the derivations that follow, we assume that agents perceive $\tilde{\eta} = \frac{\sigma_f^2}{\sigma_i^2}$ while the actual relative variance is $\eta$.\footnote{We also found that the market assumed a higher persistence. It is easy to show that this has no effect on the unconditional delayed overshooting. To save on notational complexity, we assume here that the market accurately perceives the speed of convergence of persistent shocks.} The empirical findings of section 2 imply $\tilde{\eta} >> \eta$. In
the following subsections we will derive the equilibrium exchange rate and compute the impulse responses to interest rate shocks.

We will denote by $\tilde{\alpha}_t$ the representative agent's estimator of $i^p_t$ conditional on the information set $I_t = \{I_{t-1}, i_t\}$, and $\bar{\eta}$ and by $\tilde{\sigma}_t^2$ the mean square error of this estimate. That is

$$\tilde{\alpha}_t = E[i^p_t | I_t, \bar{\eta}], \quad \tilde{\sigma}_t^2 = E[(\alpha_t - i^p_t)^2 | I_t, \bar{\eta}] \quad (2.14)$$

The corresponding estimates with the correct relative variance are denoted $\alpha_t$ and $\sigma_t^2$ respectively. During each period $t$, young investors learn from old investors that $i^p_{t-1}$ is normally distributed with mean $\bar{\alpha}_{t-1}$ and variance $\bar{\sigma}_{t-1}^2$. This induces the prior belief that $i^p_t$ is normally distributed with mean $\lambda \bar{\alpha}_{t-1}$ and variance $\lambda^2 \bar{\sigma}_{t-1}^2 + \bar{\sigma}_t^2$. After observing $i_t$ the young use Bayesian inference to update their belief about the mean of $i^p_t$. The posterior distribution of $i^p_t$ is given in the following Lemma:

**Lemma 2.1** If at time $t$ the young learn from the old that $i^p_{t-1} | I_{t-1} \sim N(\bar{\alpha}_{t-1}, \bar{\sigma}_{t-1}^2)$, then after observing $i_t$ they conclude that $i^p_t$ is normally distributed with mean $\tilde{\alpha}_t$ and variance $\tilde{\sigma}_t^2$, where:

$$\tilde{\alpha}_t = [1 - \bar{k}_t] \lambda \bar{\alpha}_{t-1} + \bar{k}_t [i_t - \bar{i}^*] \quad (2.15)$$

$$\tilde{\sigma}_t^2 = [1 - \bar{k}_t] [\lambda^2 \bar{\sigma}_{t-1}^2 + \bar{\sigma}_t^2] \quad (2.16)$$

$$\bar{k}_t = \frac{\lambda^2 \bar{\sigma}_{t-1}^2 + \bar{\sigma}_t^2}{\lambda^2 \bar{\sigma}_{t-1}^2 + \bar{\sigma}_t^2 + \bar{\sigma}_v^2} \leq 1$$

See the appendix for a proof. The $t$-young will pass on to the $t+1$-young the knowledge that $i^p_t | I_t \sim N(\tilde{\alpha}_t, \tilde{\sigma}_t^2)$. After observing $i_{t+1}$ the $t+1$-young will in turn conclude that $i^p_{t+1} | I_{t+1} \sim N(\tilde{\alpha}_{t+1}, \tilde{\sigma}_{t+1}^2)$, and pass on this information to the $t+2$-young. Therefore,
the system (2.16) can be used as a recursive updating formula under the assumption that at the beginning of history the prior was that $i_t$ was normally distributed.

The gain $\tilde{k}_t$ measures how much weight is given to new observations, which depends on the perceived quality of the public signal ($\tilde{\eta}$). No weight is given to past beliefs when there are no transitory disturbances ($\tilde{\sigma}_{x_t}^2 = 0$), as the interest rate change is perceived to be permanent. As is standard in normal updating, the gain $\tilde{k}_t$ and the variance do not depend on the particular realization of the interest rate: the system (2.16) is independent of $i_t$ and deterministic. In particular, it converges to a steady state $(\tilde{k}, \tilde{\sigma}^2)$.\(^{31}\) In what follows, we assume that the process has been going on long enough to have reached its steady state. Solving for the gain and the steady state variance of the belief, we get:

**Lemma 2.2** The steady state gain and variance are given by:

$$
\tilde{k} = \tilde{k}(\lambda, \tilde{\eta}) = \frac{1 + \tilde{\Delta} - \tilde{\eta} (1 - \lambda^2)}{1 + \tilde{\Delta} + \tilde{\eta} (1 + \lambda^2)}, \quad (2.17)
$$

$$
\tilde{\sigma}^2 = \frac{(1 - \tilde{k}) \tilde{\sigma}_x^2}{1 - (1 - \tilde{k}) \lambda^2}
$$

$$
\frac{\partial \tilde{k}(\lambda, \tilde{\eta})}{\partial \lambda} \geq 0; \quad \frac{\partial \tilde{k}(\lambda, \tilde{\eta})}{\partial \tilde{\eta}} \leq 0
$$

$$
\lim_{\tilde{\eta} \to \infty} \tilde{k}(\lambda, \tilde{\eta}) = 0; \quad \tilde{k}(\lambda, 0) = 1;
$$

where $\tilde{\Delta} = [\tilde{\eta} (1 - \lambda^2) + 1]^2 + 4 \tilde{\eta} \lambda^2$;

See the appendix. It follows from (2.17) that the gain depends only on the perceived relative variances of the noise and signal components ($\tilde{\eta}$) and the speed of convergence ($\lambda$). The gain is zero and no learning occurs when the noise is infinite while learning is immediate when the signal is perfect. The gain decreases monotonically with the noise to signal ratio. Intuitively, with a higher $\lambda$ (slower convergence), today's interest

rates contain more information about the persistent component of interest rates. As a consequence of the lemma, there is a unique \( \tilde{\eta} \) associated with any couple \( (\lambda, \tilde{k}) \) : \( \tilde{\eta} = \tilde{\eta}(\lambda, \tilde{k}) \). We can thus indifferently analyze the properties of the system in terms of \( (\lambda, \tilde{\eta}) \) or in terms of \( (\lambda, \tilde{k}) \). In the rest of this section, we parameterize our economy in terms of \( \lambda \) and \( \tilde{k} \).

In the special case where agents use the correct parameters, \( \tilde{\eta} = \eta \), we can derive the same formula, without the “tilde”. That is \( \tilde{\alpha}_t = \alpha_t, \tilde{\sigma}^2 = \sigma^2, \tilde{k} = k \) where

\[
\begin{align*}
\alpha_t &= (1 - k) \lambda \alpha_{t-1} + k [i_t - i^*] \\
k &= \frac{1 + \Delta - \eta (1 - \lambda^2)}{1 + \Delta + \eta (1 + \lambda^2)}
\end{align*}
\tag{2.18}
\]

### 2.3.2 Market Equilibrium

During each period \( f_t \) adjusts so that the market for domestic bonds clears. We assume that the population of investors has measure one. Thus, the market clearing condition is

\[
\tilde{X} = x_t (i_t, f_t, E(f_{t+1} | I_t, \tilde{\eta}))
\tag{2.19}
\]

where \( \tilde{X} \) is the aggregate supply of the domestic bond, which is fixed.

In principle, solving this problem could be quite complex since the current price \( f_t \) depends on the future price of the bond expected by each investor. To find an equilibrium we use the standard method in Finance\(^{32}\). We conjecture a linear price rule which does not convey any information beyond what is already contained in interest rates. We then solve the learning and the portfolio problems, and obtain the price function that equilibrates the domestic bond market. Lastly, we validate the initial conjecture. Using this method, we are able to exhibit one equilibrium. No claim can be made regarding uniqueness.

\(^{32}\)see Wang (1994) for an application to trading volume.

61
We make the following conjecture:

**Conjecture 2.1** The price of the domestic bond is linear in the current interest rate and in the market belief:

\[ f_t = a + b \hat{\alpha}_t + c i_t \quad (2.20) \]

The constants \( a, b \) and \( c \) will be endogenously determined. Since at time \( t \) all agents observe \( i_t \), the price does not reveal any new information. Therefore, under the price rule (2.20), the information set is \( I_t = \{ I_{t-1}, i_t, f_t \} = \{ I_{t-1}, i_t \} \).

We will use the conjectured price function to derive the optimal demand for domestic bonds by the representative young investor. We will then use the market clearing condition (2.19) to determine the market clearing price \( f_t \). As a third step, we will confirm that the market clearing \( f_t \) has the form conjectured in (2.20), and we will determine \( a, b \) and \( c \).

**Lemma 2.3** If the price function has the form conjectured in (2.20), the demand for the domestic bond is given by:

\[ x_t(f_t, i_t, \hat{\alpha}_t) = \frac{-f_t[1 + i^*] + i_t + (b + c)\lambda \hat{\alpha}_t + ci^* + a}{\gamma \pi^2} \quad (2.21) \]

where \( \pi^2 = (b^2 + c)^2 (\lambda^2 \sigma^2 + \sigma_{\varepsilon}^2 + \sigma^2) \)

The proof is in the appendix. The term \( \pi^2 \) is the conditional variance of next period's price. One can rewrite the demand function in a more familiar form:

\[ x_t(f_t, i_t, \hat{\alpha}_t) = \frac{i_t + E[f_{t+1} | I_t, \hat{\eta}] - f_t[1 + i^*]}{\gamma \pi^2} \]

The demand is increasing in the expected future price and interest rate, and decreasing in the current price and variance. Substituting (2.21) in the market clearing condition

62
(2.19) it follows that:

\[ f_t = \left[ -\gamma \pi^2 \bar{X} + c \alpha^* + a + (b + c)\lambda \alpha_t + i_t \right] [1 + i^*]^{-1} \]  \hspace{1cm} (2.22)

Lastly, equalizing the coefficients of the conjectured price function (2.20) with those of (2.22) we obtain the equilibrium price function:

**Proposition 2.1 (Equilibrium Exchange Rate)** The domestic bond's equilibrium price function is linear in the state variables and is given by:

\[ f(i_t, \alpha_t) = \frac{1 + i_t}{1 + i^*} + \frac{\lambda \alpha_t}{[1 + i^*][1 + i^* - \lambda]} - \rho \]  \hspace{1cm} (2.23)

\[ \rho = \frac{\gamma \pi^2 \bar{X}}{i^*} \]

This validates Conjecture 2.1 and characterizes the equilibrium. The equilibrium exchange rate \(1/f_t\) appreciates if the current interest rate \(i_t\) rises, or if the market belief about the persistent component of the interest rate, \(\alpha_t\), increases. It depreciates if the risk premium \(\rho\) goes up; \(\rho\) increases with the total conditional volatility of next period price \(\pi^2\), the supply of the risky asset \(\bar{X}\), and the coefficient of risk aversion \(\gamma\). Note that when the two shocks are purely transitory \((\lambda = 0)\), the market belief does not enter the pricing equation, since it does not provide any information regarding future realizations of the interest rate. It should be clear that in the case where agents use \(\eta\) the formula for the exchange rate is obtained by simply replacing \((\alpha_t, \kappa)\) by \((\alpha_t, \kappa)\).

### 2.3.3 Conditional Delayed Overshooting

Eichenbaum and Evans (1995) find that expansionary shocks to US monetary policy are followed by sharp and persistent reductions in US interest rates. Furthermore, they find
that the exchange rate systematically follows a "delayed overshooting" path: after an
initial depreciation, the US dollar continues to depreciate for several months before it
starts appreciating. A similar humped pattern is found by Clarida and Gali (1994) and
Grilli and Roubini (1994). We replicate these findings in Figure 1a-d.

The delayed overshooting path is characterized by Eichenbaum and Evans (1995) as
the impulse response of the exchange rate to an unanticipated monetary shock. In the
context of our model, we compute the path that the exchange rate would follow if a shock
to the domestic interest rate were to take place at time 0, followed by no other shocks.
In order to derive the intuition, in this subsection we consider a path conditional on a
persistent shock. In the next subsection we consider a path which is not conditioned on
whether the shock is persistent or transitory. That is, in this subsection we assume that
$\epsilon_0 = \kappa$, $\nu_0 = 0$ and $\epsilon_t = \nu_t = 0$ for $t > 0$. It follows from (2.11) and (2.12) that the
domestic interest rate follows the conditional path:

$$i_t = i^* + \lambda^t \kappa \quad \text{for all } t \geq 0 \quad (2.24)$$

We also assume for notational simplicity that in the period before the shock took
place the interest rate and the expected value of the persistent component of interest
rates were: $i_{-1} = i^*$ and $\alpha_{-1}=0$ respectively. Thus, the exchange rate $1/ f(i_{-1}, \alpha_{-1})$
was equal to $1/ \bar{f} = 1/[1 - \rho]$. 

**Full-Information Dynamics.**

It is illustrative to consider first the Full Information exchange rate path, which corre-
sponds to the Dornbusch (1976) experiment. Full information in our set-up refers to a
situation in which agents observe the current realization of the persistent component of
interest rates $i^p_t$ but ignore future shocks. It is straightforward to check, replacing $\alpha_t$ by
$i^p_t$ and adopting the same methodology as before, that the full information price is given

64
by:

\[
f^{FI}(i_t, i_t^p) = \frac{1 + i_t}{1 + i^*} + \frac{\lambda i_t^p}{[1 + i^*][1 + i^* - \lambda]} - \rho^{FI}
\]  

(2.25)

where the superscript \(FI\) stands for full information and \(\rho^{FI} = \gamma \hat{X} \left\{ \sigma_n^2 (b + c)^2 + \sigma_v^2 c^2 \right\} i^{*-1}\).

It follows from (2.25) and (2.24) that the price dynamics generated by a persistent shock under full information are given by:

\[
f_t^{FI,p} - \bar{f}^{FI} = \frac{\kappa \lambda^t}{1 + i^* - \lambda}
\]  

(2.26)

which exhibits the standard overshooting result as expected. The exchange rate appreciates initially, and from then on depreciates gradually (see ??).

**Imperfect Information Dynamics**

We assume in this section that agents form correct assessments about the relative share of transitory and persistent components: \(\hat{\eta} = \eta\) and \(\hat{k} = k\). In the case of imperfect information, by substituting (2.24) in (2.23) it follows that the initial jump in \(f\) is:

\[
f_0^\delta - \bar{f} = \delta \kappa > 0,
\]  

(2.27)

where \(\delta = \frac{1 + i^* - \lambda (1 - k)}{(1 + i^*) (1 + i^* - \lambda)} > 0\)

Thus, the exchange rate also appreciates upon impact. However, comparing (2.27) and (2.26) we can see that under imperfect information the extent of initial exchange rate appreciation is lower than under full information. Algebraically:

\[
\frac{[f_0^\delta - \bar{f}]}{[f_0^{FI,p} - \bar{f}^{FI}]} = 1 - \frac{\lambda (1 - k)}{1 + i^*} < 1.
\]
Investors are less willing to buy up the domestic asset, as the shock might prove transient. Subsequently the change in the price of the domestic bond under imperfect information is given by the following lemma, which is proven in the appendix:

**Lemma 2.4** The price change following a once and for all persistent shock is:

\[
f_{t+1}^p - f_t^p = \frac{\kappa \lambda^{t+1}}{(1 + i^*)(1 + i^* - \lambda)} \left[ (1 - k)^{t+1} (1 - \lambda (1 - k)) - (1 + i^*) \left( \frac{1}{\lambda} - 1 \right) \right]
\]  

(2.28)

Unlike in the full information case, delayed overshooting may occur under imperfect information. That is, the price of the domestic bond may increase until some time \( \tau_f > 0 \), and decline thereafter. We can determine under which circumstances this occurs by looking at the sign of (2.28). Since \( \lambda \) and \( k \) take values on \([0, 1]\) and the bracketed term in (2.28) is decreasing in \( t \), we have the following result, which is proven in the appendix:

**Proposition 2.2 (Conditional D-O Region)** A necessary and sufficient condition for delayed overshooting after \( \tau \) periods, conditional on a permanent shock at time 0, is:

\[
(1 - k)^{\tau+1} (1 - \lambda (1 - k)) - \left( \frac{1}{\lambda} - 1 \right) (1 + i^*) > 0
\]

(2.29)

This defines a delayed overshooting region

\[ D_\tau = \{(k, \lambda) | \text{such that (2.29) is satisfied}\}\]

The boundary of \( D_\tau \) is given by:

\[
\lambda(k, \tau) = \frac{(1 + i^*) + (1 - k)^{\tau+1} - \sqrt{\phi}}{2 (1 - k)^{\tau+2}}
\]

(2.30)

where \( \phi = \left[(1 + i^*) + (1 - k)^{\tau+1}\right]^2 - 4 (1 + i^*) (1 - k)^{\tau+2} \)
Note that the stylized fact that agents attach more importance to transitory shocks \((\tilde{\eta} > \eta)\) is not necessary nor sufficient for conditional delayed overshooting.

Figure 5 shows the path of the inverse of the exchange rate in response to a unit standard error increase in the interest rate at \(t = 0\) in both the perfect and imperfect information cases.\(^{33}\) It shows that a positive and persistent interest rate shock induces an initial appreciation of the exchange rate, followed by an appreciating path which lasts about twenty periods before reverting to a depreciating path. If one interprets each period as a week or a month, this graph resembles the impulse response functions estimated by Clarida and Gali (1994), Eichenbaum and Evans (1995) and Grilli and Roubini (1994). The duration of each period should depend on the frequency with which one believes that investors receive “new and relevant” information.

We now describe the intuition behind this result. There are two effects:

- **Interest rate effect.** This is the standard mechanism analyzed by Dornbusch (1976). After an initial upward jump, domestic interest rates follow a declining path. This induces the exchange rate to experience an immediate appreciation followed by a gradual depreciation to ensure that uncovered interest parity holds.

- **Learning effect.** When the shock takes place at time 0 agents only observe an increase in \(i_0\). Since the actual sequence of interest rates that agents observe (i.e. those generated by (2.24)) is greater than \(i^*\), market participants gradually increase their belief about \(i^*_t\) (i.e. \(\alpha_t\)) using updating equation (2.16). As \(\alpha_t\) is revised upwards the demand schedule for domestic bonds shifts upwards over time, generating appreciating pressures on the exchange rate.

The learning effect counteracts the interest rate effect. Therefore, upon impact the exchange rate does not jump as much as it does under perfect information. Afterwards,

\(^{33}\) The parameters chosen are: \(\lambda = 0.98, \ k = 0.2, \ i^* = 0.05\).
if the learning effect dominates, the exchange rate will continue appreciating after its initial jump as shown in Figure 5. Since agents know that \( i^* \) declines at rate \( \lambda \), at some point in time (call it \( \tau_\alpha \)) they must start revising \( \alpha \) downwards as shown in Figure 6. Thus, the exchange rate cannot continue appreciating forever. It is clear that after \( \tau_\alpha \) the exchange rate must depreciate because the learning and interest rate effects point in the same direction. Moreover, it can be shown that the exchange rate starts depreciating \textit{before} time \( \tau_\alpha \). To see this we compute from (2.16) the revisions in beliefs:

\[
\alpha_{t+1} - \alpha_t = \epsilon_0 \lambda^t \left[ (1 - k)^{t+1} (1 - \lambda (1 - k)) - (1 - \lambda) \right] \tag{2.31}
\]

from which it follows that the switching time for market beliefs is

\[
\tau_\alpha = \frac{\ln((1 - \lambda) / (1 - \lambda (1 - k))))}{\ln(1 - k)}
\]

It follows from (2.28) that the switching time for the exchange rate is

\[
\tau_f = \frac{\ln((1 - \lambda) (1 + i^*) / (1 - \lambda (1 - k)) \lambda)}{\ln(1 - k)}
\]

It is straightforward to check that \( \tau_\alpha > \tau_f \), so that \textit{the exchange rate always peaks before beliefs have started reverting}. This reflects the forward-looking behavior of the exchange rate.

Next we analyze the joint restrictions imposed by Proposition 2.2 on \( \lambda \) and \( k \) in order to deliver delayed overshooting. A smaller \( \lambda \) (less persistence) increases the second term in (2.29) proportionally more than the first one, making delayed overshooting less likely. This means that an economy converging more rapidly to its long run equilibrium is less likely to exhibit delayed overshooting. As convergence occurs faster, persistent shocks look like transitory ones. Thus, little weight is given to past observations, weakening the learning effect.
Changes in $k$ (the learning rate) have more complex effects: the first term in (2.29) is concave in $k$. For a sufficiently large $k$, the learning process works efficiently and at the time of the shock beliefs almost converge to the true value of the persistent component of the interest rate. As a consequence the subsequent upward revision of beliefs is very small. Therefore, the learning effect is dominated by the interest rate effect and there is no delayed overshooting. In other words, since beliefs have almost converged at time 0, market participants bid the exchange rate down until it is back on the full information rational expectations path. For sufficiently small $k$, learning occurs very slowly and interest rates convey little information about their persistent component. Thus, the market belief $\alpha$ increases very little at the time of the shock. Afterwards, although $\alpha$ is updated upwards, the learning effect is too small to dominate the interest rate effect.

To summarize, the simple representative agent model we consider delivers dynamics that rationalize the Eichenbaum and Evans puzzle in the case of a persistent shock. An economy will exhibit conditional delayed overshooting if the interest rate converges slowly (large $\lambda$) to its new long run value following a permanent shock to the money supply, and if learning occurs moderately slowly (intermediate values of $k$).

We now turn to the issue of the length of time over which the exchange rate moves in the “wrong” direction. As we increase the peak date $\tau$, the conditions on $\lambda$ and $k$ become more stringent: the frontier of $D_r$ shifts up, as can be seen by computing $\frac{\partial \lambda(k, \tau)}{\partial \tau}$. As shown in Figure 7a -- c, the delayed overshooting region shrinks as we increase the peak date.

Thus, our analysis has strong cross-sectional implications. Countries should exhibit delayed overshooting if (a) monetary shocks are persistent, resulting, for instance, from a low interest elasticity of money demand, and (b) the learning speed is sufficiently small, but not too small. This in turns implies a variance ratio (transitory over persistent) that is greater than one and bounded. Returning to the estimates from section 2, the first condition is likely to be satisfied, as the long run autocorrelation is usually extremely
high, indicating a high persistence, while the second is unlikely to be met: without transitory component (\( \eta = 0 \)), the inference problem is trivial and there should be no delayed overshooting.

### 2.3.4 Unconditional Delayed Overshooting

We turn in this subsection to the unconditional response of exchange rates to interest rate shocks. Eichenbaum and Evans (1995), although controlling for the source of the disturbance, do not control for the persistence of shocks. Therefore, the empirical puzzle will be solved only if one is able to generate unconditional delayed overshooting. The unconditional exchange rate path is the path following a change in the observed interest rate \( i_0 = \kappa \), not just in the persistent component \( \epsilon_0 \). We show in the appendix that given the linearity of the exchange rate equation (2.23), the unconditional path is a weighted average of the impulse responses to a persistent and a transitory shock. The weights are given by the respective conditional probabilities of both shocks. Define \( f_t^u(\kappa) \) as the exchange rate response at time \( t \) to an interest shock \( \kappa \) at time 0, where \( l = \{u, p, \tau\} \). (\( u \) stands for unconditional, \( p \) for permanent and \( \tau \) for transitory):

\[
\begin{align*}
  f_t^u(\kappa) &= E[f_t \mid i_0 - i^* = \kappa] = q f_t^p(\kappa) + (1 - q) f_t^\tau(\kappa), \\
  q &= E[\epsilon_0 \mid i_0 - i^* = 1] = \frac{\sigma^2_\zeta}{\sigma^2_\zeta + \sigma^2_\epsilon + \lambda^2 \sigma^2}
\end{align*}
\]

where \( q \kappa \) is the expected value of the persistent shock conditional on the realization of \( i_0 \). We show in the appendix that the unconditional path satisfies:

\[
\begin{align*}
  f_{t+1}^u(\kappa) - f_t^u(\kappa) &= \frac{\kappa \lambda^{t+1}}{(1 + i^*)(1 + i^* - \lambda)} \\
  &\quad \left[ (q - k)(1 - k)^t (1 - \lambda(1 - k)) - q(1 + i^*) \left( \frac{1}{\lambda} - 1 \right) \right]
\end{align*}
\]
We can sign this expression by looking at the brackets. Since $\lambda < 1$, the second term inside the brackets is negative. The first term in brackets can be positive or negative according to the sign of $q - k$. Thus, we have the following result:

**Lemma 2.5 (No Unconditional D-O)** When $\tilde{\eta} = \eta$, unconditional delayed overshooting never occurs.

The proof simply shows that $q < k$. See the appendix. The intuition for this result is straightforward: agents only observe the sum of two shocks. Conditional on this information, the persistent component of interest rates is normally distributed, with mean $q\kappa$. Hence the observed interest rate shock is equally consistent with large and positive persistent shocks $q\kappa + \epsilon$ and small or even negative persistent shocks $q\kappa - \epsilon$. In the former case, agents underreact, as in the previous subsection, while in the latter case agents initially over-react to the change in fundamentals. On average, a rational agent will not make mistakes and the standard overshooting result will apply.

In the more general case where $\tilde{\eta} \neq \eta$, there is some scope for unconditional delayed overshooting. All the calculations are the same as in the previous case replacing $\alpha_t$ by $\tilde{\alpha}_t$ and $k$ by $\tilde{k}$. Thus the unconditional path satisfies (2.34)

$$f_{t+1}^{u}(\kappa) - f_t^{u}(\kappa) = \frac{\kappa\lambda^{t+1}}{(1+i^*)(1+i^* - \lambda)} \left[ (q - \tilde{k}) \left( 1 - \tilde{k} \right)^t \left( 1 - \lambda \left( 1 - \tilde{k} \right) \right) - q (1+i^*) \left( \frac{1}{\lambda} - 1 \right) \right]$$

As in the previous case the value of $q$ is computed using the actual distribution: $q = E [\epsilon_0 | i_0 - i^* = 1, \eta] = \frac{1 - (1-\tilde{k})\lambda^2}{1+\eta(1-(1-k)\lambda)}$. Therefore, a necessary condition for unconditional delayed overshooting is $q > \tilde{k}$. Since $k$ (resp. $\tilde{k}$) is decreasing in $\eta$ (resp. $\tilde{\eta}$), we have the following result.

There are two observationally equivalent interpretations of $\tilde{\eta} > \eta$. One is that market participants make systematic errors regarding the variances of the shocks. The other is
that agents know the correct distribution, yet that the sample period under study exhibits an unusually large number of persistent shocks. The latter interpretation indicates a small sample problem over the period studied (86-95) or, equivalently, a the existence of a Peso problem. The former interpretation is consistent with a learning framework where agents learn both about the size of persistent and transitory components and their variance. However, if the parameters driving the interest rates differential process are fixed and yet unknown, the learning process should converge towards the true estimates, and expectational errors should disappear, as in Lewis (1989a). One needs a framework with stochastic regimes shifts. We plan to investigate further such a framework in the future.

Proposition 2.3 (Unconditional D-O Region) A necessary condition for unconditional delayed overshooting is that shocks are perceived as more transitory i.e., \( \hat{\eta} > \eta \).

For a given \( \eta \) a necessary and sufficient condition for unconditional delayed overshooting after \( \tau \) periods, is:

\[
(q - \bar{k}) (1 - \bar{k})^\tau (1 - \lambda (1 - \bar{k})) - \left( \frac{1}{\lambda} - 1 \right) (1 + i^*) > 0
\]  

(2.35)

If it exists, the unconditional delayed overshooting region, given \( \eta \), is defined by

\[
D_{\tau, \eta}^u = \left\{ (\bar{k}, \lambda) \mid \text{such that (2.35) is satisfied} \right\}
\]

The boundary of \( D_{\tau, \eta}^u \) is given by:

\[
\lambda(\bar{k}, \tau) = \frac{q (1 + i^*) + (q - \bar{k}) (1 - \bar{k})^\tau - \sqrt{\phi}}{2 (1 - \bar{k})^{\tau+1} (q - \bar{k})}
\]  

(2.36)

where \( \phi = \left[ q (1 + i^*) + (q - \bar{k}) (1 - \bar{k})^\tau \right]^2 - 4 (1 + i^*) (1 - \bar{k})^{\tau+1} (q - \bar{k}) \)
Figure 8a – c reports the boundaries of the unconditional delayed overshooting region when $\eta = 0.5$. It appears that the delayed overshooting region is truncated on the right: large values of the gain are not achievable anymore. For a given $\lambda$ this is equivalent to having a high $\tilde{\eta}$. The larger the true noise to signal ratio, the tighter the constraints on the delayed overshooting region. In the limit, when $\tilde{\eta}$ approaches the true $\eta$, the delayed overshooting region vanishes. Conversely, it is immediate to check that the limit of the unconditional delayed overshooting region when $\eta \to 0$ is simply the conditional delayed overshooting region, since $q \to 1$. This confirms that the most favorable case for delayed overshooting occurs when the actual shocks are largely persistent ($\eta$ close to 0) while they are perceived as transitory ($\tilde{\eta}$ large).

2.3.5 Predictable Excess Returns

In our model the uncovered interest parity condition is $E_t [f_{t+1} | I_t] + \iota_t = (1 + i^*) f_t$. To see why, note that one unit of the domestic bond has an expected return of $E_t [f_{t+1} | I_t] + \iota_t$ in terms of the consumption good. Alternatively the investor can buy $f_t$ units of the foreign bond and obtain a safe return of $1 + i^*$. We can therefore define predictable excess returns as:

$$\zeta_t = E_t [f_{t+1} | I_t] + \iota_t - (1 + i^*) f_t$$

(2.37)

In the case where $\eta = \tilde{\eta}$, substituting the equilibrium price equation (2.23) into (2.37) and taking expectations, we get:

$$\zeta_t = i^* \rho$$

(2.38)

Thus, if $\eta = \tilde{\eta}$, predictable excess returns are constant and equal to the risk premium. That is, the forward discount puzzle remains unexplained. This result implies that delayed overshooting conditional on a persistent shock does not imply time-varying predictable expected excess returns.

We now turn to the case where $\eta < \tilde{\eta}$. Substituting $\tilde{\alpha}_t$ for $\alpha_t$ in the equilibrium
price equation (2.23) and taking the expectation with respect to the actual distribution of shocks in the sample, it follows that (2.37) becomes

\[
\tilde{\zeta}_t = i^* \rho + \frac{\lambda \left( 1 + i^* - \lambda \left( 1 - \tilde{k} \right) \right)}{(1 + i^*) \left( 1 + i^* - \lambda \right)} \left( \alpha_t - \bar{\alpha}_t \right)
\] (2.39)

Predictable excess returns are time varying and are correlated, through the market belief. Under the conditions that ensure unconditional delayed overshooting, agents in the market will have systematically downward biased estimate of the persistent component. This in turn implies that the excess returns are positive and correlated through time. In other words, the predictable excess returns are positively correlated with the bond price forecast error.

These results show that a simple model of exchange rate pricing, can generate delayed overshooting and rationalize the forward discount puzzle if (a) agents are learning about the duration of interest rates shocks, and (b) the ex post distribution of shocks exhibits more persistent shocks than expected by the market. The estimates of the parameters driving the interest rates differential process, and their market perception, as reported in Tables 2.10-2.18 suggest that for most countries against the US, the conditions for delayed overshooting, according to Proposition 2.3, are satisfied.

### 2.4 An Extended Model

In the learning model of section 3, the gain and variance of the filter evolve deterministically from their initial values to their steady state values (see (2.16)). As a result, current observations of the interest rate only lead to a revision of the belief regarding the persistent component \( \alpha_t = E \left[ i_t^P \mid i_s; \ s \leq t \right] \). The linearity of the updating formula and the fact that only the first moment depends on realizations of the interest rate are responsible for the absence of predictable excess returns.

In this section we will extend the model by considering a more realistic interest rate
process in which the variance of the transitory shock is time-varying. As in section 2 the interest rate is given by (2.11) and (2.12): \( i_t = i^* + \nu_t + \epsilon_t \) and \( \nu_t = \lambda \nu_{t-1} + \epsilon_t \), where the shocks \( \epsilon_t \) and \( \nu_t \) are independent. As before, the persistent shock \( \epsilon_t \) is normal with mean zero and constant variance \( \sigma^2_{\epsilon} \). However, for the transitory shock we assume:

\[
v_t | I_{t-1} \sim N(0, \sigma^2_{\nu}(t))
\]

\[
E \left[ \nu_t^2 | \nu_{t-1} \right] = h_t^2 = \psi_0 + \psi_1 \nu_{t-1}^2;
\]

(2.40)

\[
\psi_0 > 0, 0 \leq \psi_1 < 1
\]

The autoregressive conditional heteroskedasticity (ARCH(1)) specification in (2.40) captures the tendency in financial data for volatility clustering, i.e., the tendency for large (small) price changes to be followed by other large (small) price changes of unpredictable sign. Thus, following a large transitory shock, there is an increase in variance which leads to a reduction in the precision of the investors' belief. In order to verify the validity of our decomposition of the interest rate into a persistent and a transitory component, in subsection 3.4 we estimate the process for interest rate differentials of G7 countries. We find that the interest rate differentials have a strong persistent component and a strong ARCH effect in the transitory component.

The ARCH model and its various extensions have been successfully applied to several financial time series (see Bollerslev, Chou and Kroner (1992) for a survey), including interest rates. In their study of varying risk premia in the term structure Engle, Lilien

---

The results would be equivalent if we assumed instead a time dependence on past variances, i.e., a GARCH process instead of an ARCH. In that case the investor would need to estimate today's variance of the transitory component \( E \left[ \nu_{t-1}^2 | I_{t-1} \right] \), which is equivalent to the expectation of today's conditional variance \( E \left[ h_{t-1}^2 | I_{t-1} \right] \). By the law of iterated expectations we have that \( E \left[ h_{t-1}^2 | I_{t-1} \right] = E \left[ E \left[ \nu_{t-1}^2 | \nu_{t-2}, I_{t-1} \right] | I_{t-1} \right] = E \left[ \nu_{t-1}^2 | I_{t-1} \right] \).
and Robins (1987) find very strong ARCH effects on the excess holding yield of six-month over three-month T-bills, using quarterly data from 1960-I to 1984-II. They also find ARCH effects using monthly data on one-month and two-month T-bills from 1953.1 to 1971.7. Grier and Perry (1993) look at quarterly interest rate surprises, measured as the difference between the one-month T-bill rate and the three-month forward rate for that period. They also find ARCH effects for the sample 1960-III through 1991-IV. This empirical evidence suggests that by adding an ARCH component to the interest rate process we introduce an important element of realism into the model.\footnote{Using weekly and monthly data Brenner, Harjes and Kroner (1994) estimate a model that nests the GARCH model and a model in which volatility is a function of the interest rate level. They find that their model outperforms the others because it does not underpredict volatility when interest rate levels are high as the GARCH model does, nor does it overpredict volatility in stable times as the levels model does.} In addition, this section also demonstrate that predictable excess returns can be present without unconditional delayed overshooting. The key insight is that predictable excess returns are generated by a varying risk premium while they were generated in the previous section through expectational errors.

## 2.4.1 The Learning Problem

In the extended model, young investors solve basically the same learning problem as in section 3. We assume in all this section that $\tilde{\eta} = \eta$. The only difference introduced by (2.40) is a change in the conditional distribution of $i_t - i^*$. We prove in the appendix the following result:

**Lemma 2.6** If at time $t$ the young learn from the old that $i_{t-1}^0 \mid I_{t-1} \sim N(\alpha_{t-1}, \sigma^2_{t-1})$ and if (2.40) holds, then after observing $i_t$ the young conclude that $i_t^0 \mid I_t \sim N(\alpha_t, \sigma^2_t)$, where:

$$\alpha_t = [1 - k_t] \lambda \alpha_{t-1} + k_t [i_t - i^*]$$
\[
\sigma_t^2 = \left[1 - k_t\right] \left[\lambda^2 \sigma_t^{-1} + \sigma_i^2\right]
\]

\[
k_t = \frac{\lambda^2 \sigma_t^{-1} + \sigma_i^2}{(\lambda^2 + \xi_1) \sigma_t^{-1} + \sigma_i^2 + \psi_0 + \psi_1 (i_t - i^* - \alpha_{t-1})^2}
\]

Since the \(t\)-young will in turn pass on to the \(t+1\)-young the knowledge that \(i_t^p | I_t \sim N(\alpha_t, \sigma_t^2)\) and so on, the system (2.41) can be used as a recursive updating formula under the assumption that the prior was that \(i_t^p\) was normally distributed at the beginning of history.

The updating formulas for \(\alpha_t\) and \(\sigma_t^2\) are the same as (2.16) in section 3. The only difference is that the gain \(k_t\) depends on last period's interest rate. This will be the key to generating predictable excess returns. The intuition is as follows. The last term in the denominator of the gain is the square of the estimate of the transitory shock \(\hat{\nu}_{t-1}\). An increase in \(i_{t-1}\) leads to a higher estimate of \(\nu_{t-1}\), hence a higher estimate of the variance of today's transitory shock, \(\hat{\nu}_t^2\). This decreases the gain of the filter and increases the variance in the prior distribution of \(i_t\). As a result less weight is given to current observations in updating beliefs. In the absence of shocks, the filter still converges to a steady state. However, one cannot use that steady state when inferring the size of the persistent component.

2.4.2 The Equilibrium Price Function

As before, we conjecture a linear price rule that does not reveal any information. However, since the filter is time-varying, we allow \(a_t\) to be time dependent:

\[
f_t = a_t + b \alpha_t + c i_t
\]

In equilibrium the time-varying coefficient will be the risk premium.
To solve the portfolio problem, we assume that the risk premium is unpriced risk. Thus, when solving the portfolio problem we replace $a_{t+1}$ by its expectation $\hat{a}_{t+1} = E[a_{t+1} | I_t]$ in the objective function (2.13). Note that since the interest rate shock affects the gain with a one-period lag, next period’s gain and filter are known as of time $t$. Thus, agents know at time $t$ the distribution of $(i_{t+1}, a_{t+1})$. Therefore, we can make the same factorization as in the proof of Lemma 2.3, and obtain a closed form solution for $f_t$. Following the same steps as in the previous section we find that:

**Proposition 2.4 (Equilibrium Exchange Rate)** The domestic bond’s equilibrium price function is linear in the state variables and is given by:

$$
\begin{align*}
    f_t &= \frac{i_t}{1 + i^*} + \frac{\lambda \alpha_t}{(1 + i^*)(1 + i^* - \lambda)} + a_t \\
    a_t &= \frac{1}{1 + i^*} - \gamma \bar{X} E_t \left[ \sum_{s=0}^{\infty} \frac{\pi_{t+s}^2}{[1 + i^*]^{1+s}} \right] \\
    \text{where} \quad \pi_t^2 &= \left( \frac{\lambda k_{t+1}}{(1 + i^* - \lambda)} + 1 \right)^2 \sigma_e^2 + (\lambda^2 + \psi_1) \sigma_i^2 + \psi_1 (i_t - i^* - \alpha_t)^2 + \psi_0 \frac{(1 + i^*)^2}{(1 + i^*)^2}
\end{align*}
$$

The coefficients on the interest rate and market belief are the same as in the previous section. The difference is that the independent term and the learning parameter $k_t$ are now time-varying.

Following an interest rate shock the conditional variance of next period’s interest rates (the numerator of $\pi_t^2$) goes up, leading to an increase in the risk premium. Simultaneously, the learning parameter $k_{t+1}$ goes down, pushing the risk premium down. However, it is straightforward to show that the risk premium unambiguously goes up. Over time, as agents update their beliefs, the risk premium returns to its equilibrium value. We call this mechanism the indirect learning effect. This effect captures an important feature of financial markets: a large shock leads investors to believe that there
is more uncertainty than previously expected, leading them to be more cautious. Over time, as interest rates return to normal levels, investors regain confidence.

2.4.3 Predictable excess returns

The indirect learning effect implies that the risk premium will vary over time. Accordingly, there is the potential for time-varying predictable excess returns. As noted previously, predictable excess returns can be defined as in equation (2.37): $\zeta_t = E[f_{t+1} \mid I_t] + i_t - (1 + i^*) f_t$. Replacing $f_{t+1}$ and $f_t$ by their respective formulas, we get:

$$\zeta_t = \gamma \bar{X} \left[ n_t^2 + \sum_{s=1}^{\infty} E \left[ \pi_{t+s}^2 \mid I_t \right] (1 + i^*)^{-s-1} \right]$$

(2.44)

Thus, time varying predictable excess returns reflect time varying risk premia. We observe time varying risk premia because agents' perceived risk changes over time depending on the sequence of shocks observed.

In order to see whether this time varying risk premium can explain the forward discount bias, we conducted the following Monte Carlo experiment. Using the estimated parameters for the Yen-Dollar exchange rate (see subsections 4.4), we simulated 10000 series of 1000 exchange rate and interest rate observations. For each replication, we performed the following standard "Fama" regression:

$$f_{t+1} - f_t = b_0 + b_1 (f^w_t - f_t) + u_t$$

where $f^w_t = f_t (1 + i^*) - i_t$ is the forward rate in the context of our model. Under Uncovered Interest Parity, $b_1 = 1$.\(^{36}\) Figure 9 reports the distribution of $b_1$. It appears that the distribution is centered on $\hat{b}_1 = 0.2997$ and a 95% confidence interval is $[-0.33, 0.93]$. Thus, we can reject the null. We conclude from this exercise that learning about the

\(^{36}\)Under strict UIP, we should also observe that $b_0 = 0$. However, this only represents a constant risk premium and we do not impose this restriction.
persistent component of the interest rate differential can explain some of the forward
discount bias. It is clear, however, that it cannot explain all of the bias. For the Yen-Dollar
exchange rate, Lewis (1994) finds that $b_1 = -2.28$ and is statistically significant. This
finding is confirmed by Figure 10, which reports $b_1$ against its standard error. As can be
seen on the graph, most points cluster above the non-significance cone. On average, the
simulated results report a positive, significant, but less than 1 coefficient.

Although the fluctuations in the risk premium associated with the precision of the
learning process can account for part of the forward discount puzzle, it is clear that
there is no unconditional delayed overshooting. The impulse response is now different
following a positive and a negative shock. After a positive interest rate shock, the
currency appreciates. This appreciation is dampened by the increase in risk premium.
Thus a delayed overshooting response may arise. Following a decrease in the interest
rates, however, the exchange rate depreciates. This initial depreciation is compounded by
the increase in risk premium and leads to an over depreciation. On average, the impulse
response does not exhibit any delayed overshooting.

2.4.4 Maximum Likelihood Estimation of the Interest Rate Process

In this subsection we confirm the validity of our assumptions regarding the decomposition
of the interest rate into a persistent and a transitory component using data for the G7
countries. Figure 11 reports the interest rate differential against the US from 1974 to
1992. If the interest rate satisfies (2.11), (2.12) and (2.40), we can write the conditional likelihood of $i_t$ as:

$$f_{i_t|I_{t-1}} (i_t|I_{t-1}) = \frac{1}{(2\pi)^{1/2} \omega_t^{-1}} \exp \left( \frac{(i_t - \lambda \alpha_{t-1})^2}{2 \omega^2_t} \right)$$

where

$$\omega^2_t = \left( \lambda^2 + \psi_1 \right) \sigma_{t-1}^2 + \sigma^2_e + \psi_0 + \psi_1 (i_{t-1} - i^* - \alpha_{t-1})^2$$

80
We can then maximize the sample log likelihood \( \sum_{t=1}^{T} \log f_{i_t|t-1} (i_t|I_{t-1}) \) with respect to the vector of parameters \( \theta = (\lambda, \psi_0, \psi_1, \sigma_\epsilon, i^*)' \). Note that this estimator is also a Method of Moments estimator corresponding to the first order condition

\[
E \left[ \frac{\partial \log f (i_t|I_{t-1}, \theta)}{\partial \theta} \right] = 0
\]

as long as the unconditional expectation of the score vector is well defined. We use monthly IFS data for G7 countries against the US, from 1974:1 to 1992:12. Our measure of short term interest rates is the short term money market rate (line 160bc) used in Grilli and Roubini (1994). In order to have a more flexible parameterization, we also allow for an AR\( (p) \) in the persistent component and an ARCH\( (q) \) in the transitory component. Thus we estimate the following model:

\[
di_t = i_t - i^*_t = i^* + \lambda(L) d_i^p_t + \nu_t; \quad \nu_t|I_{t-1} \sim N(0, \sigma_\nu^2(t))
\]

\[
\lambda(L) d_i^p_t = \epsilon_t; \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)
\]

\[
E \left[ \nu_t^2 | \nu_{t-1} \right] = h^2_t = \psi_0 + \psi(L) \nu_{t-1}^2;
\]

\[
\psi_0 \geq 0, \psi_t \geq 0, \psi(1) < 1
\]

where \( \lambda \) is a polynomial in the lag operator of order \( p \) and \( \psi \) is of order \( q \). The constant \( i^* \) captures the non zero mean of the interest rate differential, and in the context of the model it can be interpreted as the average constant return.

The estimation procedure is similar to section 2. In practice, increasing the order of the ARCH effect did not improve the results substantially. In contrast, increasing the order of the AR effect was crucial for correctly identifying the model. For low AR orders, the estimation procedure fails to exhibit significant transitory components (both the ARCH coefficients and \( \psi_0 \) are extremely small). We therefore report the results for
<table>
<thead>
<tr>
<th>U.S. vs</th>
<th>U.K.</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Canada</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>2.0372</td>
<td>2.3206</td>
<td>2.2374</td>
<td>1.7405</td>
<td>1.5693</td>
<td>2.0295</td>
</tr>
<tr>
<td>(0.4849)</td>
<td>(0.1848)</td>
<td>(0.1150)</td>
<td>(0.3522)</td>
<td>(0.4216)</td>
<td>(0.2376)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-1.4913</td>
<td>-1.9897</td>
<td>-1.9817</td>
<td>-1.1336</td>
<td>-1.0777</td>
<td>-1.5128</td>
</tr>
<tr>
<td>(0.8172)</td>
<td>(0.3278)</td>
<td>(0.2018)</td>
<td>(0.5704)</td>
<td>(0.5846)</td>
<td>(0.4191)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>0.4282</td>
<td>0.6516</td>
<td>0.7388</td>
<td>0.3472</td>
<td>0.4385</td>
<td>0.4638</td>
</tr>
<tr>
<td>(0.3699)</td>
<td>(0.1609)</td>
<td>(0.1059)</td>
<td>(0.2414)</td>
<td>(0.2493)</td>
<td>(0.2035)</td>
<td></td>
</tr>
<tr>
<td>$\psi_0$</td>
<td>0.8082</td>
<td>0.1516</td>
<td>0.1550</td>
<td>0.0552</td>
<td>0.3252</td>
<td>0.0658</td>
</tr>
<tr>
<td>(0.5170)</td>
<td>(0.1002)</td>
<td>(0.1430)</td>
<td>(0.1544)</td>
<td>(0.2725)</td>
<td>(0.0448)</td>
<td></td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>(0.3983)</td>
<td>(0.2405)</td>
<td>(0.3864)</td>
<td>(0.6129)</td>
<td>(0.2453)</td>
<td>(0.1673)</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.5967</td>
<td>0.1313</td>
<td>0.2022</td>
<td>0.9878</td>
<td>0.7858</td>
<td>0.4242</td>
</tr>
<tr>
<td>(0.5997)</td>
<td>(0.0813)</td>
<td>(0.0847)</td>
<td>(0.5348)</td>
<td>(0.6240)</td>
<td>(0.2074)</td>
<td></td>
</tr>
<tr>
<td>$t^*$</td>
<td>-1.5698</td>
<td>-1.5262</td>
<td>1.6352</td>
<td>-5.7499</td>
<td>-1.8215</td>
<td>1.9401</td>
</tr>
<tr>
<td>(4.8681)</td>
<td>(2.3156)</td>
<td>(5.6382)</td>
<td>(2.5369)</td>
<td>(1.6417)</td>
<td>(3.3419)</td>
<td></td>
</tr>
<tr>
<td>log L</td>
<td>-0.6014</td>
<td>-0.3104</td>
<td>-0.3352</td>
<td>-0.3657</td>
<td>-0.4823</td>
<td>-0.2967</td>
</tr>
<tr>
<td>$\Psi(1)$</td>
<td>0.9740</td>
<td>0.9825</td>
<td>0.9945</td>
<td>0.9540</td>
<td>0.9301</td>
<td>0.9806</td>
</tr>
<tr>
<td>(0.3983)</td>
<td>(0.2405)</td>
<td>(0.3864)</td>
<td>(0.6129)</td>
<td>(0.2453)</td>
<td>(0.1673)</td>
<td></td>
</tr>
</tbody>
</table>

Source: IFS.monthly money market rates. Sample: 1974:1-1992:12. Estimates equation 2.45 by maximum likelihood. Arch coefficients are constrained to be positive and $\Phi_1 \leq 1$. Standard errors are reported in parentheses. The updating equations are reported in the appendix for the general case AR(p), ARCH(q). Results for an AR(3) ARCH(3) were very similar and are available upon request.
an AR(3), ARCH(1) specification in Table 2.19.

The results indicate strong persistence in the interest differential. Indeed, for most countries, the sum of the AR coefficients is close to 0.95 and strongly significant. In addition, the results indicate a very strong ARCH effect. In all cases, $\psi_1 = 1$. We plot in Figure 12 the persistent and transitory innovations to the interest rate differentials. One can see that the increase in volatility is associated with the 1979-1985 period and the associated change in the Fed's operating procedure. Interestingly, the volatility of the transitory component of the estimation procedure is strong for the US-UK interest rate differential. Comparatively, shocks to the US-Japan interest rate differential are much weaker. Table 2.19 reflects the comparatively larger $\psi_0$ estimated for these two countries. The US-Italian interest rate differential is the only one exhibiting a higher volatility of the persistent component.

2.5 Conclusion

We have presented a model of nominal exchange rate determination that exhibits the delayed overshooting pattern of exchange rates found by Eichenbaum and Evans (1995). Conditional delayed overshooting results from the interaction of learning about the current state of affairs and the intrinsic dynamic response of interest rates to monetary shocks. This interpretation, which is new to our knowledge, has important implications. First, it provides a clear analytical characterization of the factors influencing exchange rate responses to monetary shocks. Countries with rapidly converging interest rates, due to either fast moving prices or a large interest elasticity of money demand, will experience less delayed overshooting. Countries with either a very small or a very large variance of transitory shocks will also converge without delayed overshooting: in the former case because learning occurs fast, in the latter case because learning does not have a significant

\[37^\text{The only exception is US-Italy.}\]
effect on the demand for assets.

Second, we have shown that a simple extension of our model can rationalize unconditional delayed overshooting and predictable excess returns. Our key assumptions is that the sample distribution and its market expectation differ, reflecting either small sample problems or expectational errors. Typically, unconditional delayed overshooting arises when shocks are more permanent than expected by market participants. We have found, using survey data on interest rate forecast that this assumption seem to be strongly supported: while the data fail to exhibit significant transitory components, market participant implicitly assume that a sizeable portion of the shocks is transitory. Moreover, estimating the model on monthly data for G7 countries, we found that our modelling assumptions accurately characterize the interest differential process and that the coefficients are often in the “delayed overshooting region.”

Lastly, we have shown that a simple extension of our model can rationalize part of the forward discount puzzle without generating unconditional delayed overshooting. We found that deviations from Uncovered Interest Parity can arise if we make the additional assumption that the quality of the learning process is affected by the size of the shocks – that is, if shocks to the interest rate increase the variance of shocks. This assumption generates time varying risk premia, which in turn induce systematic deviations from uncovered interest parity.
Appendix A

a  Delayed Overshooting

We replicated the results of Eichenbaum and Evans (1995) using a slightly longer time period. The data are monthly, from the IFS, and the sample period is 1974:1 to 1992:12. Eichenbaum and Evans covers 1974:1 to 1990:5 while Grilli and Roubini covers 1974:1 to 1991:12. The interest rates are monthly market rates. The exchange rates are quoted as units of foreign currency per dollar. The chapter estimates recursive VARS. See Kim and Roubini (1995) and Clarida and Gali (1994) for structural VAR, identified with contemporaneous and long run restrictions respectively. The countries in our sample are: UK, France, Germany, Italy, Canada and Japan.

Eichenbaum and Evans (1995) consider three measures of monetary policy: the ratio of Non Borrowed Reserves to total reserves, the Federal Funds rate, and the Romer and Romer index of monetary policy. They look at the response of exchange rates to innovations in their measure of monetary policy. We replicated two of their specifications:

- \( \{Y_{us}, CPI_{us}, NBRX_{us}, R^* - R_{us}, E\} \). In this specification, the innovations to the ratio of Non Borrowed Reserves represent monetary shocks. The Fed observes domestic industrial production and prices. All variables are in level form. Figure 1a-b show the impulse response of the interest rate differential (top row) and the nominal exchange rate (bottom row) for 8 G7 countries. In all cases, the path of interest differential is consistent with a Dornbusch-type experiment. The exchange rate response exhibits substantial inertia (especially in the cases of Japan, Canada and the UK). We report standard deviation bands around point estimates which were computed using a Monte Carlo method with 500
Table A.1: ARMA, Euro 3 Months

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
<td>0.934</td>
<td>0.855</td>
<td>1.418</td>
<td>0.738</td>
<td>0.892</td>
<td>1.150</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>0.035</td>
<td>(0.151)</td>
<td>(0.103)</td>
<td>(0.042)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>-0.427</td>
<td>0.358</td>
<td>-0.213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.123)</td>
<td>(0.068)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>-0.155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_1 )</td>
<td>-0.238</td>
<td>0.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.132)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_2 )</td>
<td>0.238</td>
<td>0.543</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.073)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>591.32</td>
<td>849.58</td>
<td>536.20</td>
<td>232.75</td>
<td>186.33</td>
<td>575.27</td>
</tr>
</tbody>
</table>


draws.

- \( \{Y_{us}, CPI_{us}, Y^*, R^*, FF, NBRX_{us}, E \} \). In this specification, the Fed also observes foreign output and interest rates before setting the federal funds rate. We plot the response of Non Borrowed Reserves to total reserves (NBRX in the top row) and the nominal exchange rate (bottom row). See Figure 1c-d. Following an increase in the Federal Funds rate, the ratio of Non Borrowed Reserves dips down and then increases over time. The nominal exchange rate exhibits a pattern closely resembling the unconditional impulse response we obtained in section 3. On impact, the exchange rate appreciates, then it depreciates rapidly, sometimes falling below its original level. After 5 to 10 periods, the delayed overshooting pattern emerges.

b  Empirical Results on Interest Rate Differentials

We describe in this section the empirical procedure. The survey data are described in section 2 in the chapter. We first run ARMA processes of various orders and select the ones that minimize Akaike's criterion. This procedure is unconstrained for both the order of the AR and MA component (the maximum order is 5 on each component). The results are reported in Table A.1, A.2 and A.3. Results using Euro-3 months interest rates report robust standard errors.
Table A.2: ARMA, Money Market Rates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.742</td>
<td>0.901</td>
<td>0.946</td>
<td>0.9103</td>
<td>0.7249</td>
<td>1.391</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.028)</td>
<td>(0.019)</td>
<td>(0.027)</td>
<td>(0.043)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\varphi_1$</td>
<td></td>
<td>-0.283</td>
<td></td>
<td>-0.385</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td></td>
<td>(0.062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>1014.65</td>
<td>690.24</td>
<td>759.92</td>
<td>689.94</td>
<td>837.11</td>
<td>612.31</td>
</tr>
</tbody>
</table>


Table A.3: ARMA, Prime Rate

<table>
<thead>
<tr>
<th></th>
<th>UK-US</th>
<th>Canada-US</th>
<th>Japan-US</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>1.171</td>
<td>0.942</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.034)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.195</td>
<td></td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>(0.100)</td>
<td></td>
<td>(0.137)</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td></td>
<td></td>
<td>-0.301</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.101)</td>
</tr>
<tr>
<td>AIC</td>
<td>138.70</td>
<td>134.98</td>
<td>105.24</td>
</tr>
</tbody>
</table>

In all specifications but UK-US and Germany-US for the euro 3 months, the order of the moving average is inferior to the order of the autoregressive part and satisfies the overidentification restrictions. In a majority of cases, there is no moving average component indicating that there is no associated transitory component.

b.1 Kalman Filter Estimation

This subsection briefly derives the Kalman Filter equations. We postulate the following process:

\[ d_{it} = H'\xi_t + \nu_t \quad (b.1) \]
\[ \xi_t = F\xi_{t-1} + \epsilon_t \quad (b.2) \]

where \( \xi_t = (d_{it}, ..., d_{i,t-p+1})' \), \( H' = (1, 0, ..., 0)' \) is a px1 vector. \( \xi_t \) is the state vector for the process, (b.1) the measurement equation and (b.2) the space equation. Define the informations set \( I_t = \{d_{i,t-i}, i \geq 0\} \), \( \hat{\xi}_{t+1|t} = E[\xi_{t+1}|I_t] \), and \( \hat{P}_{t+1|t} = E \left[ (\xi_{t+1} - \hat{\xi}_{t+1|t}) (\xi_{t+1} - \hat{\xi}_{t+1|t})' |I_t \right] \).

The filtering equations are:

\[ \hat{\xi}_{t+1|t} = F\hat{\xi}_{t-1|t} + \hat{F}\hat{\xi}_{t-1|t}H' \left( H'\hat{P}_{t|t-1}H + \sigma^2 \right)^{-1} \left( \xi_{t+1|t} - H'\hat{\xi}_{t|t-1} \right) \]
\[ \hat{P}_{t+1|t} = F \left( \hat{P}_{t|t-1} - \hat{P}_{t|t-1}H' \left( H'\hat{P}_{t|t-1}H + \sigma^2 \right)^{-1} H'\hat{P}_{t|t-1} \right) F' + \sigma^2 \]

The smoother equations are:

\[ \hat{\xi}_{t|T} = \hat{\xi}_{t|t} + \hat{F}\hat{\xi}_{t+1|T} \left( \hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t} \right) \]
\[ \hat{P}_{t|T} = \hat{P}_{t|t} + \left( \hat{P}_{t|t}F'\hat{P}_{t+1|t}^{-1} \right) \left( \hat{P}_{t+1|T} - \hat{P}_{t+1|t} \right) \left( \hat{P}_{t|t}F'\hat{P}_{t+1|t}^{-1} \right)' \]

where

\[ \hat{P}_{t|t} = \hat{P}_{t|t-1} - \hat{P}_{t|t-1}H' \left( H'\hat{P}_{t|t-1}H + \sigma^2 \right)^{-1} H'\hat{P}_{t|t-1} \]
c Proofs.

Proof of Lemma 2.1:
In order to prove this Lemma we will use a standard result in Bayesian inference (see DeGroot (1970, section 9.5)). Suppose that $y_t$ is an observation from a normal sampling distribution with unknown mean $\theta$ and known variance $\rho^2$. If the prior distribution of $\theta$ is normal with mean $m_{t-1}$ and variance $\tau_{t-1}^2$, then the posterior distribution of $\theta$ after observing $y_t$ is normal with mean $m_t$ and variance $\tau_t^2$, where

$$m_t = (1 - k_t)m_{t-1} + k_t y_t, \quad \tau_t^2 = (1 - k_t)\tau_{t-1}^2, \quad k_t = \frac{\tau_{t-1}^2}{\tau_{t-1}^2 + \rho^2} \quad (c.3)$$

To apply this result to our model note that the observation $y_t$ corresponds to $i_t - i^*$ and that $\theta$ corresponds to $i_t$, and recall the notation $E \left[ i_t^p \mid I_{t-1} \right] = \bar{\alpha}_{t-1}$ and $E \left[ (\bar{\alpha}_{t-1} - i_t^p)^2 \mid I_{t-1} \right] = \bar{\sigma}_{t-1}^2$. First, since the young investors know that $i_{t-1}^p \mid I_{t-1} \sim N(\bar{\alpha}_{t-1}, \bar{\sigma}_{t-1}^2)$, and since $i_t^p = \lambda i_{t-1}^p + \epsilon_t$, their prior about $i_t^p$ (before observing $i_t$) is $i_t^p \mid I_{t-1} \sim N(\bar{\alpha}_{t|t-1}, \bar{\sigma}_{t|t-1}^2)$, where

$$\bar{\alpha}_{t|t-1} = E \left[ i_t^p \mid I_{t-1} \right] = \lambda \bar{\alpha}_{t-1} \quad (c.4)$$

$$\bar{\sigma}_{t|t-1}^2 = E \left[ (\bar{\alpha}_{t|t-1} - i_t^p)^2 \mid I_{t-1} \right] = E \left[ (\lambda \bar{\alpha}_{t-1} - \lambda i_{t-1}^p - \epsilon_t)^2 \mid I_{t-1} \right] = \lambda^2 \bar{\sigma}_{t-1}^2 + \bar{\sigma}_t^2$$

Second, note that the observation $i_t - i^* = i_t^p + \nu_t$ is normal with unknown mean $i_t^p$ and known variance $\bar{\sigma}_t^2$. Lastly, using the above result it follows that after observing $i_t$ the young investors’ posterior is $i_t^p \mid I_t \sim N(\bar{\alpha}_t, \bar{\sigma}_t^2)$, where $\bar{\alpha}_t$ and $\bar{\sigma}_t^2$ are given by (2.16). The equations in (2.16) are obtained by substituting $(i_t - i^*, \lambda \bar{\alpha}_{t-1}, \bar{\alpha}_t, \bar{\sigma}_{t|t-1}^2, \bar{\sigma}_t^2, \bar{\sigma}_t^2)$ for $(y_t, \mu_{t-1}, \mu_t, \tau_{t-1}^2, \tau_t^2, \rho^2)$ in (c.3).

Proof of Lemma 2.2
To prove this lemma, we simply solve for the constant gain and variance. Dividing both sides by $\bar{\sigma}_t^2$ we find that the gain and variance only depend on the noise to signal ratio. The derivatives are straightforward.

Proof of Lemma 2.3
Using the conjectured price function to eliminate $f_{t+1}$ from the objective function, we get:

$$U (c_t, c_{t+1}) = -\exp (-\gamma c_t) \quad (c.5)$$

89
\[\delta \exp \left( -\gamma \left[ W_t - f_t x_t - c_t \right][1 + i^*] - \gamma x_t [a + i_t] \right) E \left[ \exp \left( -\gamma x_t B' \theta_{t+1} \right) \right] \]

where \(B' = (b, c)\) and \(\theta'_{t+1} = (\tilde{a}_{t+1}, i_{t+1})\). The vector \(\theta_{t+1}\) is normally distributed with mean \(\tilde{\theta}_{t+1}\) and covariance matrix \(\Sigma\)

\[\tilde{\theta}_{t+1} \quad \sum (\lambda^2 \sigma^2 + \sigma^2 \sigma^2) \begin{bmatrix} \bar{k}^2 & \bar{k} \\ \bar{k} & 1 \end{bmatrix} \]

Therefore, the expectation in (c.5) is given by

\[E \left[ \exp \left( -\gamma x_t B' \theta_{t+1} \right) \right] = \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \theta_{t+1} \Sigma^{-1} (\theta_{t+1} - \tilde{\theta}_{t+1}) \right) \frac{1}{\sqrt{2\pi}} \\text{d} \theta_{t+1} \]

where \(m = \tilde{\theta}_{t+1} - \Sigma \gamma x_t B\). The second equality follows from the fact that covariance matrix \(\Sigma\) is symmetric. Since the integrand in the second equation in (c.7) is the density function of a normal variable, it follows that the integral is equal to one. Thus, the expectation in (c.5) is equal to

\[\exp \left( -\gamma x_t B' \tilde{\theta}_{t+1} + \frac{1}{2} \gamma^2 x_t^2 B' \Sigma B \right) \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \theta_{t+1} \Sigma^{-1} (\theta_{t+1} - m) \right) \frac{1}{\sqrt{2\pi}} \\text{d} \theta_{t+1} \]

The second equality is obtained by carrying out the summation and setting \(\alpha_{-1} = 0\). Equation (2.28) is obtained by using (c.8) to compute \(f_{t+1}^p - f_t^p\) and by setting \(\epsilon_0 = \kappa\).

**Derivation of Proposition 2.2:**

The first part of the proposition is trivial. To find the frontier of \(D_r\), we set (2.29) equal to zero. This gives a second order equation in \(\lambda\). We select the root smaller than 1, which is given by (2.30).

**Derivation of Equations (2.32) and (2.33)**

Denote by \(f_t(\epsilon, \nu)\) the impulse response at time \(t\) to an initial persistent shock of size \(\epsilon\) accom-
panied by a transitory shock of size $\nu$. The unconditional path (2.32) is given by

$$
\begin{align*}
f_t^\kappa &= \int f_t(\epsilon_0, \kappa - \epsilon_0) \, d\varphi(\epsilon_0|\epsilon_0 = \kappa) = \int \left( f_t(\epsilon_0, 0) + f_t(0, \kappa - \epsilon_0) \right) \, d\varphi(\epsilon_0|\epsilon_0 = \kappa) \\
&= f_t^\kappa(\kappa) \, E[\epsilon_0|\epsilon_0 = 1] + f_t^\kappa(\kappa) \, E[\epsilon_0|\epsilon_0 = 1]
\end{align*}
\tag{c.9}
$$

The second and third equalities follow from the linearity of the impulse response in the original shocks. To derive the expression for $E[\epsilon_0|\epsilon_0 = \kappa]$ in (2.32) we use formula (c.3). Let us consider $i_0 - i^* = \lambda \epsilon_0^{(0)} + \nu_0 + \epsilon_0$ as an observation from a distribution parameterized by $\epsilon_0$. Note that since the system was in steady state (i.e., the expectation of $i_0^{(0)}$ was zero), it follows that $i_0 - i^* \sim N(\epsilon_0, \lambda^2 \sigma^2 + \sigma^2_\epsilon)$. Note also that the prior distribution of $\epsilon_0$ is normal with mean zero and variance $\sigma^2_\epsilon$. Thus, substituting $(i_0 - i^*, 0, E[\epsilon_0|i_0 - i^*], \sigma^2_\epsilon, \text{var}[\epsilon_0|i_0 - i^*], \lambda^2 \sigma^2 + \sigma^2_\epsilon)$ for $(y_t, \mu_{t-1}, \mu_t, \tau_t^2, \tau^2_t, \rho_t)$ in (c.3), it follows that the posterior distribution of $\epsilon_0$ is $\epsilon_0 | i_0 - i^* \sim N\left(\frac{\sigma^2_\epsilon(i_0 - i^*)}{\lambda^2 \sigma^2 + \sigma^2_\epsilon} + \frac{\sigma^2_\epsilon^2}{\lambda^2 \sigma^2 + \sigma^2_\epsilon}, \frac{\sigma^2_\epsilon}{\lambda^2 \sigma^2 + \sigma^2_\epsilon}\right)$. By setting $i_0 - i^* = \kappa$ we obtain (2.32). To obtain (2.33) note that the impulse response to a persistent shock is given by (c.8). Substituting $i_0 = i^* + \kappa, i_t = i^*$ for $t > 0$ and $\alpha_t = \lambda^t (1 - k)^t k \kappa$ in (2.23), it follows that the impulse response to a transitory shock of size $\kappa$ is:

$$
f_t^\kappa(\kappa) = \tilde{f} + \frac{\lambda^{t+1} (1 - k)^t k \kappa}{[1 + i^*][1 + i^* - \lambda]} \tag{c.10}
$$

By substituting (c.8) and (c.10) in (c.9), and taking first differences we obtain (2.33).

**Proof of Lemma 2.6:**

Since $q = \frac{\sigma^2_\epsilon}{\lambda^2 \sigma^2 + \sigma^2_\epsilon} < \frac{\sigma^2_\epsilon}{\sigma^2_\epsilon + \sigma^2_\kappa}$, it is sufficient to show that $\frac{1}{1 + \eta} < k$. Using the definition of $k$, in (2.17), we get

$$
\frac{1}{1 + \eta} < k = \frac{1 + \Delta - \eta (1 - \lambda^2)}{1 + \Delta + \eta (1 + \lambda^2)}
$$

By rearranging it follows that this inequality holds if and only if $1 + \eta (1 - \lambda^2) < \Delta$. Lastly, the definition of $\Delta$ in (2.17) implies that this inequality always holds. Thus, $q < k$.

**Proof of Lemma 2.6:**

The problem is the same as that of Lemma 2.1. The only difference is the distribution of $i_t - i^*$. We will show that $i_t - i^*$ is normally distributed with unknown mean $i^*_t$ and known variance.
given by
\[ E \left[ \nu^2_{t-1} \mid I_{t-1} \right] = \xi_0 + \xi_1 \left[ (i_{t-1} - i^* - \alpha_{t-1})^2 + \sigma^2_{t-1} \right] \]  \hspace{1cm} (c.11)

Recall that \( \alpha_{t-1} = [i^p_{t-1} \mid I_{t-1}] \) and \( \sigma^2_{t-1} = E \left[ (i^p_{t-1} - \alpha_{t-1})^2 \mid I_{t-1} \right] \). To derive (c.11) note first that from (2.40) and the law of iterated expectations it follows that
\[ E \left[ \nu^2_{t} \mid I_{t-1} \right] = E \left[ E \left[ \nu^2_{t} \mid \nu_{t-1}, I_{t-1} \right] \mid I_{t-1} \right] = E \left[ h^2_{t} \mid I_{t-1} \right] = \xi_0 + \xi_1 E \left[ \nu^2_{t-1} \mid I_{t-1} \right] \]  \hspace{1cm} (c.12)

Second, to derive \( E \left[ \nu^2_{t-1} \mid I_{t-1} \right] \) note that
\[ E \left[ \nu^2_{t-1} \mid I_{t-1} \right] = E \left[ (i_{t-1} - i^* - i^p_{t-1})^2 \mid I_{t-1} \right] \]  \hspace{1cm} (c.13)
\[ = E \left[ (i_{t-1} - i^*)^2 - 2 (i_{t-1} - i^*) i^p_{t-1} + (i^p_{t-1})^2 \mid I_{t-1} \right] \]
\[ \sigma^2_{t-1} = E \left[ (i^p_{t-1} - \alpha_{t-1})^2 \mid I_{t-1} \right] = E \left[ (i^p_{t-1})^2 \mid I_{t-1} \right] - \alpha^2_{t-1} \]  \hspace{1cm} (c.14)

Using (c.14) to eliminate \( E \left[ (i^p_{t-1})^2 \mid I_{t-1} \right] \) from (c.13) and substituting the result in (c.12) we obtain (c.11). Next, following the same steps as in the proof of Lemma 1 and using the same notation as in (c.4), we have that the prior distribution of \( i^p_{t-1} \) is normal with mean \( \alpha_{t|t-1} = \lambda \alpha_{t-1} \) and variance \( \sigma^2_{t|t-1} = \lambda^2 \sigma^2_{t-1} + \sigma^2 \). Lastly, to obtain the mean and variance of the posterior distribution of \( i^p_{t} \) we substitute \( (i_1 - i^*, \alpha_{t|t-1}, \alpha_t, \sigma^2_{t|t-1}, \sigma^2_t, E \left[ \nu^2_{t} \mid I_{t-1} \right]) \) for \( (y_t, \mu_{t-1}, \mu_t, \tau^2_t, \tau^2_t, \rho^2) \) in (c.3). The resulting expressions are given by (2.41).
Figure 2.1a: Impulse Responses to Monetary Innovations (ratio of non-borrowed reserves)
Figure 2.1.b: IMPULSE RESPONSES TO MONETARY INNOVATIONS (RATIO OF NON-BORROWED RESERVES)
Figure 2.1.d: Impulse Responses to Monetary Innovations (Federal Fund Rate)
Figure 2.2: MONEY MARKET AND PRIME RATE
Figure 2.3: Money Market and Euro 3-months Rate
Figure 2.4: EURO 3 MONTHS AND FORECASTS
Figure 5: Delayed Overshooting

$\varepsilon_0 = 1; \lambda = 0.98; k = 0.2; i^* = 0.05$
Figure 2.6: Market Belief

Figure 6: Market Belief

$\varepsilon_0 = 1, \lambda = 0.98, k = 0.2, i^* = 0.05$
Figure 2.7.a: **Conditional Delayed Overshooting Region**

Figure 7.a: Delayed Overshooting Region

$\tau = 1; \varepsilon_0 = 1; \theta = 0.05$
Figure 2.7.b: CONDITIONAL DELAYED OVERTSHOOTING REGION
Figure 2.7.c: CONDITIONAL DELAYED OVERSEASHOOTING REGION

Figure 7.c: Delayed Overshooting Region

\( \tau = 10; \quad \varepsilon_0 = 1; \quad \gamma = 0.05 \)
Figure 2.8.b: UNCONDITIONAL DELAYED OVERTSHOOTING REGION

Figure 8.2: Unconditional Delayed Overshooting Region

$\epsilon_0 = 1; \ i^* = 0.05; \ \eta = 0.5$

$k$

$\chi$
Figure 8.c: Unconditional Delayed Overshooting Region

\[ \tau = 10; \quad \varepsilon_0 = 1; \quad i^* = 0.05; \quad \eta = 0.5 \]
Figure 2.9: MONTE CARLO DISTRIBUTION OF $\beta_1$
Figure 2.10: MONTE CARLO SIMULATION: $\beta_1$ vs $se(\beta_1)$
Figure 2.12: INNOVATIONS TO DIFFERENTIALS

Figure 12A: Interest Rate Differentials
UK-US

Figure 12B: Interest Rate Differentials
FR-US

Figure 12C: Interest Rate Differentials
GE-US

Figure 12D: Interest Rate Differentials
IT-US

Figure 12E: Interest Rate Differentials
CA-US

Figure 12F: Interest Rate Differentials
JA-US
Chapter 3

Non-Linearities in Exchange Rates: A Non-Parametric Instrumental Variable Approach

1 Introduction

The literature on exchange rate target-zone models has exploded in the past decade, following Krugman's (1991) seminal contribution. Thanks to the simplicity and elegance of Itô's calculus, economists have been able to explore analytically the implications of exchange rates bands on depreciation expectations and validate Williamson's (1985) analysis. These models typically predict some form of non-linearity between the "fundamental" and the nominal exchange rate. Moreover, these non-linearities exert stabilizing effects on the exchange rate. This is an area of international economics where data are abundant and well calibrated: exchange rates are available at high frequency, almost without any measurement error. In addition, understanding the behavior of exchange rate in presence of bands is of paramount importance for many European policymak-
ers. Thus, the theory has been the subject of numerous empirical investigations. Given the inherent unobservability of the fundamental, empirical work has taken two different routes: the first one consists in testing the model indirectly using implications for observable variables such as the interest rate differential or the exchange rate itself. The second tests directly for non-linear effects by constructing a proxy for the fundamental. Empirical results following both approaches have been disappointing for the basic target zone theory. In particular, Flood, Rose and Mathieson (1991) using direct estimation, claim that there are no significant non-linearities.

In this chapter I argue that the second approach -direct estimation- faces serious econometric problems. More specifically, due to measurement errors, it fails the orthogonality condition. As the chapter demonstrates, constructed fundamentals tend to be extremely correlated with the exchange rate itself. This in turn implies that the estimators might be biased against finding any non-linear effect. Measurement error in linear models, once identified, is easily corrected by using additional information. Two-stage least square yields a consistent estimator when observations on instrumental variables are available. Measurement error in non-linear models, unfortunately, cannot be cured so easily. In particular, Instrumental Variables estimators are not consistent. Intuitively, the error-in-variable does not “add-up” linearly with the true error term of the regression. This breaks down the equivalence between measurement error and simultaneous equations models. The functional form is, in some specific sense, convoluted with the error term. Conditions under which one can estimate parameters consistently are usually extremely specific and depend on the particular functional form of the problem or properties of the data generating process.

In order to solve this problem, I adopt a non-parametric approach that allows me to instrument the fundamental while preserving consistency. The main idea consists in relaxing the functional specification imposed by target zones models, in favor of a richer and more versatile structure. A Non-Parametric Instrumental Variable (henceforth
NPIV) estimator based on Series estimation is then presented. Under some regularity conditions, this estimator will be consistent. This estimator has the flavor of getting something for nothing. Unfortunately, this is not the case, as I will not be able anymore to test the parametric implications of target-zone theory. Instead I am forced to adopt a somewhat "agnostic" position whereby the focus is more on generic non-linearities in exchange rates than on the specifics of target zones models. However, since there are a variety of refinements of the basic theory, each of which implies a slightly different form of non-linearity, this is also an advantage of my procedure.

Moreover, the NPIV estimator is general enough to be applied to a range of problems focusing on non-linearities, where exogenous variables suffer from measurement problems. Examples include, and are not limited to, investment dynamics, or finance models of stock prices behavior.\textsuperscript{1} Obviously, this method requires the identification of potential instruments. In the case of exchange rate models, I argue that the excess return differential, the excess of stock returns over bonds in the domestic versus the foreign country, constitutes such an instrument. It is uncorrelated with the currency risk premium, the source of measurement error in our context, and is likely to be correlated with the "fundamentals" when the latter reflect the state of the economy, broadly defined.

Applying this method to EMS countries since 1979 on daily data, I find preliminary evidence of substantial non-linearities. Moreover, these non-linearities appear to have been stabilizing at both edges of the band, for a number of countries. These results, in sharp contrast with earlier studies, highlight the dangers of measurement error for non-linear models. Exchange rate non-linearities are the rule, not the exception. Thus our estimation procedure delivers dramatically different results than the previous literature.

Section 2 briefly summarizes the standard results on target zones models, emphasizing the empirically testable implications of the theory. Readers interested mostly

\textsuperscript{1}See Caballero and Engel (1994) for a different approach aimed at solving a similar problem in the investment literature.
in the definition of the \textit{NPIV} estimator should go directly to section 3. I apply this methodology to exchange rates in section 4, and compares the results to traditional, but problematic, results before concluding.

2 Exchange rate target zones, a brief review of the theory and empirical results.

2.1 The prototype model;

The simplest target zones model of exchange rate can be summarized by the following two equations:\footnote{See Krugman (1991).}

\[
e_t = k_t^* + \alpha E_t (d_e_t) / dt
\]  
\[
k_t^* = \eta dt + \sigma dz
\]

\(e_t\) represents the log of the nominal exchange rate and \(k_t^*\) the "fundamental" driving process. Equation (2.1) usually represents the reduced form of a more structural model. In the monetary rendition, equation (2.1) derives from both 	extit{Uncovered Interest Parity} (UIP) and 	extit{Purchasing Power Parity} (PPP). The fundamental is then a function of money supply and output.\footnote{For instance, a simple symmetric monetary model gives \(k_t^* = (m_t - m_t^*) - \beta (y_t - y_t^*)\) where \(m_t\) and \(y_t\) represent respectively domestic money supply and output and foreign variables are starred.} Less stringent versions allow a role for sticky prices by introducing a Phillips curve type equation in place of PPP. Appropriately defined, the fundamental then depends also on the real interest rate and its expected change.\footnote{see Flood and Rose (1993)} In both versions, \(\alpha\) represents the interest semi-elasticity of money demand.

More generally, one could interpret equation (2.1) as a pricing equation. The exchange rate today is determined by "fundamental" factors and by the conditional expectation of the exchange rate tomorrow. However, (2.1) alone is not enough to solve the model.
We need a distributional assumption on the fundamental $k^*_t$. Equation (2.2) states that $k^*_t$ follows a standard Brownian motion with drift $\eta$ and variance $\sigma^2$, between two interventions by the monetary authorities.

In a free float regime (no intervention), we find the solution of (2.1) by integrating forward, assuming non bubble:

$$e_t = k^*_t + \alpha\eta$$

(2.3)

Without interventions, the expected rate of depreciation of the currency is the same as the rate of depreciation of the fundamental, $\eta$. The nominal exchange rate varies linearly with the fundamental.

By contrast, in a target-zones regime, authorities announce that they will allow the exchange rate to fluctuate freely within some pre-defined band $[e_l, e_u]$. When the exchange rate reaches its boundaries, the Central Bank takes whatever action is needed to prevent any move outside the band. This raises two issues: one is the credibility of the central bank announcement, the other concerns the means of defending the band. In this simple version it is assumed that the announcement is fully credible (so that the exchange rate is never expected to move outside the bands), and the central bank adjusts marginally, shifting the fundamental by infinitesimal amounts.\(^5\) In technical terms, the fundamental follows a regulated Brownian motion:\(^6\)

$$dk_t = \eta dt + \sigma dz + dU - dL$$

(2.4)

where $dU$ and $dL$ are non-decreasing discontinuous processes, which adjust upwards only when the exchange rate reaches its lower (respectively upper) bound. In this regime, expectation of exchange rate depreciation are no longer constant: intuitively, when the exchange rate moves closer to the upper boundary, say, agents can safely expect an

---

\(^5\)As opposed to infra-marginal adjustments occurring in the interior of the band or discrete adjustments that would shift discretely the exchange rate and the fundamental back in the interior of the band.

\(^6\)See Harrison (1985).
appreciation, as the central bank is likely to intervene. By (2.1), this implies that the exchange rate today is lower. Formally, noting that the fundamental follows a strong Markov process and that the exchange rate is a function only of the current fundamental, one can apply Itô’s lemma to (2.1) and find the general solution:

\[ e_t = k_t^* + \alpha \eta + A_1 \exp(\lambda_1 k_t^*) + A_2 \exp(\lambda_2 k_t^*) \]  \hspace{1cm} (2.5)

where

\[ \lambda_{1,2} = -\eta \pm \sqrt{\eta^2 + 2\sigma^2/\alpha} \geq 0 \]  \hspace{1cm} (2.6)

and \( A_1 \) and \( A_2 \) are integration constants satisfying the following boundary conditions, together with the boundaries of the fundamental process \( k_u, k_l \):

\[
\begin{align*}
\{ & e_u = k_u + \alpha \eta + A_1 \exp(\lambda_1 k_u) + A_2 \exp(\lambda_2 k_u) \\
& e_l = k_l + \alpha \eta + A_1 \exp(\lambda_1 k_l) + A_2 \exp(\lambda_2 k_l) \\
& 0 = 1 + A_1 \lambda_1 \exp(\lambda_1 k_u) + A_2 \lambda_2 \exp(\lambda_2 k_u) \\
& 0 = 1 + A_1 \lambda_1 \exp(\lambda_1 k_l) + A_2 \lambda_2 \exp(\lambda_2 k_l)
\end{align*}
\]  \hspace{1cm} (2.7)

(2.7) is known as the Value Matching condition. It imposes continuity at the boundaries, while (2.8) is a Smooth Pasting (or super-contact) condition, implying that there is no-kink in \( de_t/dk_t^* \).\footnote{See Dumas (1991) for a discussion of these conditions.}

With these boundary conditions, one can show that \( A_1 < 0 \) and \( A_2 > 0 \): the relationship between \( e_t \) and \( k_t \) exhibits the familiar S-shape, characteristic of the target zone literature and reproduced on Figure 1. The exchange rate fluctuates less in a target zones regimes (solid line) than under a free float (dashed line). In other words, speculation exerts a stabilizing effect, sometimes dubbed the honeymoon effect.\footnote{Note also that the function relating the exchange rate and the fundamental is single valued in the case of marginal infinitesimal adjustments.}

118
This stabilization effect relies crucially on the sign of $A_1$ and $A_2$ and gives an empirically testable restriction of the theory, if one could observe the “fundamental”.

2.2 Extensions;

The bare-bone model has strong empirical implications. Before turning to the empirical literature, it is worth investigating whether similar implications hold true when one relaxes the key assumptions of perfect credibility and marginal, infinitesimal adjustments.

Imperfect credibility;

When the Central Bank’s announcements are not completely credible, the exchange rate is not expected to stay within the band with probability 1. A collapse could occur, with reversion to free float as in the literature on speculative attacks, or the monetary authorities could decide to defend a new band, by a discrete jump in the central parity. In the first case, the stabilizing effect will be weaker but still present, as the exchange rate will—at most—jump back on the free float line. In the second case, however, as shown by Bertola and Caballero (1992), the target zone may have a destabilizing effect, if the expected jump in the central parity is large enough. Whenever the exchange rate moves closer to the upper boundary, agents now expect a large realignment that will shift the exchange rate above the free float line. Thus the exchange rate today depreciates even further. This is obviously a limit case, driven by specific assumptions on the realignment process.

Discrete interventions;

The general solution to (2.1) is still given by (2.5) in the presence of discrete marginal interventions. Bertola and Caballero (1992) prove that the integration constants $A_1$ and $A_2$ are indeed unchanged. Discrete interventions only affects $k_u$ and $k_l$, the barriers of the fundamental’s process. As a result, the relationship between the exchange rate and
the fundamental is not one to one any more. Stabilizing effects are still present but may be harder to detect.

**Inframarginal adjustments and multilateral systems;**

The case of continuous inframarginal adjustments was first analyzed by Delgado and Dumas (1990), for a bilateral target zone. These authors simply assumed that $k$ follows a mean-reverting Ohrstein-Uhlenbeck process. The stabilizing effects are still present, yet largely explained by the managed float component (the mean reverting part of the Ohrstein Uhlenbeck process) rather than the presence of the target zone. As a result, the non-linear terms in (2.5) are smaller.

A recent literature has investigated the implications of inframarginal interventions in the context of a multilateral target zone. In a multilateral target zone, the exchange rate between currency 1 and 2, $e_{1,2}$, may reach its limits while the exchange rate between currency 1 and 3, $e_{1,3}$, is still strictly inside its band. As long as monetary authorities in country 1 intervene to bring $e_{1,2}$ strictly inside its band, the fundamental $k_{1,3}$ is likely to be indirectly affected.\(^9\) This has two implications. First, there will be an inframarginal adjustment on $e_{1,3}$. Second, and more importantly, $e_{1,3}$ may not be able to reach its "notional" boundaries since some other currency would reach its limits first. This leads to the notion of "effective" bands, which may be everywhere tighter than "notional" bands Pill (1994). In an $N$-country target-zone, any bilateral exchange rate will be a function of the $N - 1$ independent fundamentals. Serrat (1994) solves this problem in closed form and shows that Krugman's solution only applies when (a) the notional and effective bands coincide everywhere, and (b) one country plays the role of an anchor and never intervenes. Moreover, Serrat (1994) shows that the "honeymoon" effect, although still present, depends on the multidimensional path of the fundamental vector. Looking

\(^9\)Strictly speaking, only one country needs to intervene to bring the currency back in the band. In our example, if country 2 adjusts and country 1 does not, there is no shock to the fundamental $k_{1,3}$. In the EMS, the largest multilateral target-zone, both countries are forced to intervene at the boundaries.
simply at the bilateral band, one might not find evidence of the standard stabilization
effect and wrongly conclude that the theory is rejected. However, it still remains the
case that the exchange rate is not a linear function of the fundamental. Therefore, on a
broad level, investigating the non-linear relationship between the fundamental and the
exchange rate remains a valid question.

Therefore, adding elements of realism to the bare-bone model may weaken -or even
reverse- the honeymoon effect, but not eliminate it altogether. That the recent models
focus on elements likely to weaken or negate the stabilizing effect is no wonder when one
considers the empirical evidence on target zones accumulated in the recent years.

In what follows, I will remain squarely within the boundaries of the bilateral target-
zone model. As already mentioned, in a multilateral target zone, any exchange rate
depends on a multidimensional fundamental vector. Although this represents no serious
difficulty from the theoretical point of view, it adds a substantial number of state vari-
ables. In a non-parametric set-up, as I will develop shortly, this worsens the "curse of
dimensionality". Identification of non-linear effects might then prove even more tenuous.

2.3 The empirical literature.

The literature on the subject is extremely varied and adopts different approaches that
yield the same negative conclusion: stabilizing non-linear effects are hard to come by
in the data. We review here the main results and methodologies. They fall into two
broad categories. The first methods try to test indirectly the model, by concentrating
on distributional implications of the model for observable variables (e.g., interest rate
differential, exchange rate). The second approach tries to estimate directly non-linear
terms in (2.5).
The ergodic distribution approach.

From the distribution assumption (2.2) and the nature of the process, it is possible to write the ergodic distribution of the fundamental $\phi(k^*)$. $k^*$ has a truncated exponential distribution in the presence of drift ($\eta \neq 0$). Without drift, the ergodic distribution is uniform. With infinitesimal interventions, $e(k^*)$ is one to one and onto. Therefore, one can make a change of variable and compute the ergodic distribution of $e_t$. Svensson (1991) shows that the ergodic distribution thus obtained is U-shaped: the exchange rate should spend most of its time close to the boundaries. The intuition for this result is as follows: when the exchange rate depreciates, the down-side risk of further depreciation is reduced, as the exchange rate is closer to the upper bound. In turn, this induces private speculative flows betting on an appreciation of the currency. This offsets the impact of positive velocity shocks. As a result, the exchange rate will exhibit a tendency to remain where it is. This mechanism is stronger, the closer the exchange rate is to its boundaries.

This implication of the theory appears grossly violated by the facts: Flood et al. (1991) and Bertola and Caballero (1992) show that the empirical frequency of the French Franc-Deutschemark exchange rate since 1979 is hump-shaped. In other words, the exchange rate tend to spend most of its time near the central parity. However, neither paper present a formal test of departure from the ergodic distribution. In addition, the theory predicts a hump-shaped ergodic distribution in the case of discrete interventions, or of multilateral target-zone Serrat (1994).

The interest rate differential.

If one is willing to assume that Uncovered Interest Parity holds, the interest differential constitutes a measure of expected exchange rate depreciation. In a credible target zone regime, the exchange rate should never be expected to move out of the band. Whenever it moves closer to the upper bound, it should be expected to appreciate, leading to a negative correlation between the interest differential and the exchange rate. This
constitutes and indirect - and joint- test of the model. The empirical evidence gathered by Svensson (1991) on EMS and Scandinavian countries rejects strongly this implication, indicating some peso effect.

Maximum Likelihood estimation.

I now turn to direct estimation procedures of the non-linear effects. Maximum Likelihood approach, developed by Pesaran and Samiei (1992b) and (1992a)estimates a discrete version of (2.1) instead of (2.5), as a limited-dependent rational expectation model. The exchange rate is assumed to follow a two-limit Tobit:

$$
\begin{align*}
&
\begin{cases}
  e_t = e_u & e_t^* \geq e_u \\
  e_t = e_t^* & e_l \leq e_t^* \leq e_u \\
  e_t = e_l & e_t^* \leq e_l
\end{cases}
\end{align*}
$$

(2.9)

where

$$
e_t^* = \alpha E_{t-1}(e_t) + X_t \beta + u_t \text{ and } X_t = z_t R + v_t
$$

(2.10)

The latent variable $e_t^*$ depends on the expected exchange rate $E_{t-1}(e_t)$ and some exogenous and potentially unobservable variables $X_t$. $X_t$ in turn is a function of some observed variables $z_t$. The authorities control the exchange rate through $u_t$. Assuming that $(u_t, u_l)$ is normally distributed, they solve first for the expected exchange rate as a function of the observed variables and parameters, then write the full-information likelihood. In this approach, non-linear effects come through the conditional expectation of the exchange rate. Pesaran and Samiei (1992a) estimate their model on the French Franc-Deutschemark exchange rate from 1979 to 1989, using money supply, output and interest differentials as exogenous variables, at a monthly frequency. Their results show that the fit of the model is greatly improved once we allow for non-linear effects. However these authors implicitly assume that the band was unchanged and totally credible over the ten years of their sample. This is an heroic assumption. If realignments occur, as
it did six times in the period under study, expectations of exchange rate depreciation are modified, leading to more complex expressions for the likelihood. As these authors recognize, their result may simply pick up the linear relationship between the exchange rate and the central parity over the sample period.

**Direct estimation of the non linear terms.**

Direct estimation of (2.5) is difficult, given that the fundamental $k_t^*$ is unobservable. It is, however, the most direct way to test for the presence of non linear terms and their potentially stabilizing effects. Flood et al. (1991) pursue this approach using high frequency data on exchange rate and interest rates for most EMS countries. Assuming that *Uncovered Interest Parity* holds:

$$i_t - i_t^* = E_t(det)/dt$$  \hspace{1cm} (2.11)

and substituting in (2.1), one can derive a proxy $k_t$ for the unobservable fundamental $k_t^*$ for different values of $\alpha$:10

$$k_t = e_t - \alpha(i_t - i_t^*)$$  \hspace{1cm} (2.12)

With this proxy in hand, one can estimate (2.5) and a discrete version of (2.2) using a two step OLS method. In the first step, they estimate the parameters of the fundamental process $\hat{\eta}$ and $\hat{\sigma}^2$ from a regression of $k_t$ onto itself lagged:

$$k_t = k_{t-1} + \eta + \epsilon_t \text{ where } \text{var}(\epsilon_t) = \sigma^2$$  \hspace{1cm} (2.13)

From this estimated values they construct $\hat{\lambda}_1$ and $\hat{\lambda}_2$ according to (2.6), and the pseudo-regressors $\exp(\hat{\lambda}_1 k)$, $\exp(\hat{\lambda}_2 k)$. They then estimate (2.5), adding an error term on the

---

10Their approach consists in using out of sample information for $\alpha$, from money demand equation, for instance. In practice, they assume $\alpha = 0.1$. 

124
right hand side and allowing for some misspecification, by OLS:

\[ e_t - k_t - \alpha \hat{\eta} = A_0 + A_1 \exp(\lambda_1 k_t) + A_2 \exp(\lambda_2 k_t) + A_3 k_t + u_t \quad (2.14) \]

The null hypothesis is:

\[ H_0 : A_1 < 0, A_2 > 0 \]

This methodology runs into several problems:

(i) Looking at (2.11), it is immediate that the constructed fundamental and the nominal exchange rate will be "approximately" the same variable. The interest rate differential fluctuates much less - at high frequency - than the exchange rate itself. This was pointed out by Lewis (1991): if we denote by \( f_d \) the forward rate defined from Covered Interest Parity, one can rewrite (2.11) as: \(^{11}\)

\[ k_t = (1 + \alpha) e_t + \alpha f_d t \quad (2.15) \]

The constructed fundamental is a weighted sum of the current exchange rate and the forward rate, both of which are extremely similar. Figure 2 presents the DM-FF log exchange rate and the constructed fundamental for the period 1/12/87-1/5/90. It is apparent that the two series are almost identical. Figure 3 also reports the interest differential between these two countries. It represents a negligible fraction of the fundamental. Thus, this methodology biases the results against non-linear terms: it is no wonder that the linear term captures almost everything. \(^{12}\)

---

\(^{11}\) Covered Interest Parity holds on Eurocurrencies as opposed to Uncovered Interest Parity.

\(^{12}\) Assume, as an extreme example, that there is no interest differential in the sample. Then \( k_t = e_t \) and a regression of \( e_t \) on \( (k_t, f_1 (k_t), f_1 (k_t)) \), where \( f_1 (.) \) and \( f_1 (.) \) are non linear functions, will find a coefficient of exactly 0 on all non-linear terms.
(ii) More importantly, this points towards an error in variable problem. The fundamental is a sufficient statistic for the exchange rate process. This method assigns as a sufficient statistic for the exchange rate...the exchange rate itself (up to the interest rate differential). Even under the very unlikely hypothesis that equation (2.5) holds exactly, Uncovered Interest Parity cannot be expected to be satisfied, even for very short horizons. The traditional biases and inconsistencies related to errors in variable problem should be expected.

Stated differently, this method fails the Orthogonality Condition and the estimators are inconsistent. In the following section, I propose a Non-Parametric Instrumental Variable designed to solve this type of problem.

3 An agnostic approach using Non-Parametric Instrumental Variables (NPIV).

3.1 Exposition of the Problem

Suppose economic theory predicts the following relationship between the exchange rate and the true "fundamental" \( k^*_t \):

\[
\epsilon_t = \bar{f}(k^*_t, \theta^0) + \epsilon_t
\]

\[
E[\epsilon_t | k^*_t] = 0,
\]

where \( \bar{f}(.,.) \) is some known parametric function, possibly non-linear, and \( \theta^0 \) is a vector of unknown parameters controlling the shape of the function. We are interested in estimating \( \theta^0 \). For instance, (2.5) implies that the exchange rate depends non-linearly on the fundamental and on the vector \( \theta = (A_1, A_2) \). Assume in addition that \( k^*_t \) is
unobserved. Instead, we observe the variable $k_t$, related to $k_t^*$ by the measurement equation:

$$k_t = k_t^* + \eta_t,$$

$$E[\eta_t | k_t^*, \epsilon_t] = 0$$

Using $k_t$ instead of $k_t^*$ will give an inconsistent estimator of $\theta^0$. Replacing $k_t^*$ by $k_t - \eta_t$ one gets:

$$e_t = f_k (k_t - \eta_t, \theta^0) + \epsilon_t$$

$$= f_k (k_t, \theta^0) + \epsilon_t - f_k (k_t, \theta^0) + \epsilon_t - \eta_t \theta^0$$

where $\tilde{k}_t$ is a stochastic variable lying between $k_t$ and $k_t^*$, and $f_k (\cdot, \cdot)$ denotes the partial derivative of $f (\cdot, \cdot)$ with respect to its first argument.\(^{(13)}\) Define the pseudo-residual $\tilde{\epsilon}_t = \epsilon_t - \tilde{f}_k (\tilde{k}_t, \theta^0)$. Direct estimation is inconsistent:

$$E [\tilde{\epsilon}_t | k_t] = -E \left[ f_k (\tilde{k}_t, \theta^0) | \eta_t + k_t^* \right] \neq 0.$$

In the linear case, $f_k (k, \theta) = k \theta$ and (3.18) simplifies to

$$e_t = k_t \theta^0 + \epsilon_t = k_t \theta^0 + \epsilon_t - \eta_t \theta^0$$

The pseudo-residual is not a function of $k_t$ anymore, yet direct estimation is still incon-

\(^{(13)}\) Implicitly I assume that the function $f$ is sufficiently regular for all relevant partial derivatives to exist and be continuous.
sistent since

\[ E[\tilde{\epsilon}_t|k_t] = -E[\eta_t|\eta_t + k_t^*] \theta^0 \neq 0. \]

The linear problem can be solved by using additional information, for instance instrumental variables. Suppose \( z_t \) is an instrument for \( k_t \), satisfying:

\[ E[\eta_t|z_t] = E[\epsilon_t|z_t] = 0; \ E[z_tk_t] = R \]

(3.19)

Then:

\[ E[e_t - k_t\theta^0|z_t] = 0 \]

which provides a basis for estimation using Indirect Least Squares.\(^{14}\) In the linear case, the measurement error problem is isomorphic to the Simultaneous Equations problem, and the same technique can be employed to solve both problems. In the non-linear case, the pseudo-residual \( \tilde{\epsilon}_t \) depends on \( k_t \). This dependance in turn breaks down the equivalence between Simultaneous Equations and measurement error problems. More precisely,

\[
E[e_t - \tilde{f}(k_t, \theta^0)|z_t] = E[\tilde{\epsilon}_t|z_t]
= -E[\tilde{f}_k(k_t, \theta^0)|z_t]
\neq 0
\]

where the last inequality comes precisely because \( k_t \) and \( z_t \) are correlated, a property required for \( z_t \) to be a valid instrument. Thus Non-Linear Instrumental Variables, as developed by Amemyia (1974) does not give a consistent estimator. Some results

---

\(^{14}\)Or Two-Stage Least Squares, if one has more instrumental variables than regressors.
might be obtained under additional moment restrictions. Amemyia (1985) shows that one can estimate consistently $\theta^0$ when the variance of the measurement error goes to 0 asymptotically.\textsuperscript{15} This is a somewhat restrictive assumption: in many economic context the covariance matrix may not be shrinking asymptotically. In the exchange rate context, as will see in the next section, the measurement error is related to the currency premium. There is no reason \textit{a priori} to expect the currency premium to be small or to become deterministic asymptotically.

3.2 Non-Parametric Instrumental Variable (NPIV)

In this section, I take a different track, extending a procedure developed by Hausman et al. (1991) for polynomial regressions to the Non-Parametric case. In particular, I will not assume that the covariance matrix of the measurement error vanishes asymptotically. In what follows, I assume that a valid instrument is available and postpone until next section the question of the choice of such an instrument in the context of exchange rate models.

The strategy consists in relaxing the tight functional specification imposed by the target zone model. Parametric models provide consistent and efficient estimators under appropriate regularity conditions. The tight jacket of parameter-vector estimation gives powerful results whenever the model is correctly specified. When the parametric model is not true, these estimates may not be efficient or consistent any more. In the context of target-zone, the parametric model represented by (2.5) imposes a very stringent structure on the data that makes difficult any attempt to estimate the model using \textit{constructed} fundamentals. Relaxing this structure for a-priori less stringent representations might improve our results. Since it ignores information contained in the theoretical functional form $\tilde{f}(.,.)$, the estimator is likely to be inefficient.

The intuition behind this approach is as follows. Assume $z_i$ is a valid instruments for

\footnote{This might be the case if a large number of measurement on each regressor is available.}
More specifically, assume that

\[ k_t^* = g(z_t) + \nu_t \]
\[ E[\nu_t | z_t] = 0 \]

where \( g(.) \) is some unspecified. The conditional expectation of \( e_t \) given \( z_t \) is given by:

\[ E[e_t | z_t] = \int \bar{f}(g(z_t) + \nu_t, \theta^0) \varphi_Z(\nu_t) \, d\nu_t \]  \quad (3.20)

where \( \varphi_Z(.) \) is the p.d.f. of \( \nu_t \) conditional on \( Z \). Equation (3.20) shows that, in general, \( E[e_t | z_t] \) will depend on the true parameters \( \theta^0 \) and the moments of \( \nu_t \). The latter are "nuisance parameters" that one would like to eliminate. The conditional expectation convolutes \( \bar{f}(\cdot, \cdot) \). In order to estimate \( \theta^0 \), one needs to "deconvolute" (3.20). This is usually extremely difficult. Suppose now that we "forget" the restrictions imposed by the theory on \( \bar{f}(\cdot, \cdot) \) and adopt the agnostic position that \( e_t \) is related to \( k_t \) through some unspecified non-parametric function \( f(.) \). The objective now, is to recover the function \( f(.) \) itself, instead of the parameter \( \theta^0 \). At first glance, it might seem that our prospects have not improved substantially with this relaxation of the problem. I will show that this is indeed not the case.

Assume the following representation for the exchange rate process:

\[ e_t = f(k_t^*) + \epsilon_t, \]  \quad (3.21)
\[ k_t = k_t^* + \eta_t, \]  \quad (3.22)
\[ k_t^* = g(z_t) + \nu_t, \]  \quad (3.23)

\[ E[\epsilon_t | z_t, \nu_t] = E[\eta_t | z_t, \nu_t, \epsilon_t] = 0; \]  \quad (3.24)
\[ E[v_t] = 0, \text{ and } v_t \text{ is independent of } z_t. \]  \hspace{1cm} (3.25)

\[ E[\epsilon_t \eta_t | z_t] = \sigma_{\epsilon \eta}; \quad E[\epsilon_t v_t | z_t] = \sigma_{\epsilon v} \]  \hspace{1cm} (3.26)

Both \( f(\cdot) \) and \( g(\cdot) \) are non-parametric functions. (3.21) states that the nominal exchange rate is a non-parametric function of the unobserved fundamental. (3.22) is the measurement equation. The observed fundamental is equal to the true fundamental plus some noise \( \eta_t \). (3.23) relates the "true" fundamental to the instruments \( z_t \). We simply assume that the fundamental is related to the instruments in a non-parametric and unknown way through the function \( g(\cdot) \).\(^{16}\)

(3.24) is necessary for \( z_t \) to be a valid instrument. (3.25) is crucial: the independence of \( v_t \) and \( z_t \) allows to recover structural parameters. Note that this is a stronger assumption than is usually needed for instrumental variables. Under regularity conditions, \( f(\cdot) \) can be expressed as a Fourier series:

\[ f(k) = \sum_{i=0}^{\infty} \beta_i p_i(k) \]  \hspace{1cm} (3.27)

where \( \{p_i(\cdot)\}_{i=0}^{\infty} \) represents a known basis of polynomial functions and \( \{\beta_i\}_{i=0}^{\infty} \) is the family of Fourier coefficients. Thanks to variants of Newton formula, one can also typically write an addition formula of the form:

\[ p_i(x + y) = \sum_{j=0}^{i} \lambda_{ij} p_j(\tau x) p_{i-j}(\tau y), \]  \hspace{1cm} (3.28)

and a recursive formula:

\[ x p_i(x) = \omega_i p_{i+1}(x) + \chi_i p_{i-1}(x) \]  \hspace{1cm} (3.29)

\(^{16}\) \( z_t \) being an instrument for \( k_t^* \), the functional form \( g(\cdot) \) is not of interest in itself: any function of \( z_t \) is also an instrument. This representation is a convenient way to write directly the optimal instrument as a combination of functions of \( z_t \). See Newey (1990).
for some coefficients \( \{ \lambda_{ji} \} \), \( \{ \omega_i \} \) and constant \( \tau \).\(^{17}\) This implies that polynomials have nice convolution properties that will make possible the deconvolution of the true functional form from the nuisance parameters. More specifically, using (3.28),

\[
E \left[ e_t \mid z_t \right] = \int f \left( g(z_t) + u_t \right) \varphi_Z(v_t) \, dv_t \\
= \int \left( \sum_{i=0}^{\infty} \beta_i p_i \left( g(z_t) + u_t \right) \right) \varphi_Z(v_t) \, dv_t \\
= \sum_{i=0}^{\infty} \beta_i \sum_{j=0}^{i} \lambda_{ji} p_j \left( \tau g(z_t) \right) \int p_{i-j} \left( \tau v_t \right) \varphi_Z(v_t) \, dv_t \\
= \sum_{j=0}^{\infty} p_j \left( \tau g(z_t) \right) \left( \sum_{i=j}^{\infty} \beta_i \lambda_{ji} E \left[ p_{i-j} \left( \tau v_t \right) \mid z_t \right] \right) \\
= \sum_{j=0}^{\infty} p_j \left( \tau g(z_t) \right) \gamma_j
\]

(3.30)

The second line uses (3.27), the third (3.28) and the last comes from reordering the double summation. The last equality highlights the convolution of the Fourier parameters (the \( \beta_i \)'s), with moments of the noise \( E[p_{i-j} \left( \tau v_t \right) \mid z_t] \) in constructing the \( \gamma \)'s. Clearly, (3.30) is not enough to estimate the Fourier coefficients. However, one can also use an auxiliary regression, proposed by Hausman et al. (1991), together with the recursive formula (3.29):

\[
E \left[ e_t k_t \mid z_t \right] = \int \left( g(z_t) + u_t + \eta_t \right) \left( f \left( g(z_t) + u_t \right) + e_t \right) \varphi_Z(v_t) \, dv_t \\
= \int \left( g(z_t) + u_t + \eta_t \right) \left( \sum_{i=0}^{\infty} \beta_i p_i \left( g(z_t) + u_t \right) + e_t \right) \varphi_Z(v_t) \, dv_t
\]

\(^{17}\)In the case of simple powers, \( \lambda_{ji} = \binom{i}{j} \), \( \omega_i = 1 \), \( \chi_i = 0 \), and \( \tau = 1 \). The addition and recursion formulas for Hermite polynomials are presented in the next subsection. We define \( p_{-1} \) so that (3.29) is valid for \( i = 0 \).
\[ \begin{align*}
&= \int (g(z_l) + v_l) \left( \sum_{i=0}^{\infty} \beta_i p_i (g(z_l) + v_l) \right) \varphi_Z (v_l) \, dv_l + \sigma_{\eta l} + \sigma_{\epsilon l} \\
&= \sum_{i=0}^{\infty} \beta_i \int (w_{i} p_{i+1} (g(z_l) + v_l) + \chi_{i} p_{i-1} (g(z_l) + v_l)) \varphi_Z (v_l) \, dv_l + \sigma_{\eta l} + \sigma_{\epsilon l} \\
&= \sum_{i=0}^{\infty} (\beta_{i-1} w_{i-1} + \beta_{i+1} \chi_{i+1}) \int p_i (g(z_l) + v_l) \varphi_Z (v_l) \, dv_l + \sigma_{\eta l} + \sigma_{\epsilon l} \\
&= \sum_{i=0}^{\infty} (\beta_{i-1} w_{i-1} + \beta_{i+1} \chi_{i+1}) \sum_{j=0}^{\infty} \lambda_j p_j (\tau g(z_l)) E [p_{i-j} (\tau v_l) | z_l] + \sigma_{\eta l} + \sigma_{\epsilon l} \\
&= \sum_{j=0}^{\infty} p_j (\tau g(z_l)) \left( \sum_{i=0}^{\infty} (\beta_{i-1} w_{i-1} + \beta_{i+1} \chi_{i+1}) \lambda_j E [p_{i-j} (\tau v_l) | z_l] \right) + \sigma_{\eta l} + \sigma_{\epsilon l} \\
&= \sum_{j=0}^{\infty} p_j (\tau g(z_l)) \delta_j \\
&= (3.31)
\end{align*} \]

where the third line uses (3.24) (3.25) and (3.26), the fourth uses (3.29) and the sixth uses (3.28). I implicitly defined \( \beta_{-1} = 0 \). Direct estimation of (3.30) and (3.31) will yield consistent estimates of both \( \gamma \)'s and \( \delta \)'s. Deconvoluting is simply a matter of inverting the formulae for \( \gamma \) and \( \delta \) in (3.30) and (3.31) and separating recursively the structural parameters from the nuisance parameters.

In practice, one cannot estimate all the Fourier coefficients, both in the reduced and structural forms. Therefore, I will select an order of approximation \( K (n) \) for \( f \) and \( L (n) \) for \( g \), and perform the estimation on this finite order polynomials approximation.\(^{18}\)

\[ f (k) = \sum_{j=0}^{K} \beta_j p_j (k); \quad g (z) = \sum_{i=0}^{L} \alpha_i p_i (z) \quad (3.32) \]

Since I can, following Hausman et al. (1991), estimate consistently the parameters of the series representation for any order \( K, L \), assuming regularity conditions on the

\(^{18}\)Hausman et al. (1991) have solved a similar problem under the assumption that \( f (.) \) is a polynomial and \( g (.) \) is linear. Under some regularity assumptions, their estimators are both consistent and asymptotically normal.
functional $f$, and $g$, I can derive a sequence of series approximation $\hat{f}_n$ that is consistent for $f$, in the sense that $\lim \hat{f}_n (k) = f (k)$ for any $k$. Non-linearities will imply significant high order Fourier coefficients. Thus, the null becomes:

$$H_0 : \beta_j = 0, \ j \geq 2$$

In order to estimate $\beta$, one must first choose a suitable family of polynomials $\{p_i(\cdot)\}$. The ultimate issue concerns the optimal choice of $K(n)$ and $L(n)$, where $n$ is the number of observations, so as to insure asymptotic convergence of the approximation.

**Normalized Hermite Polynomials**

Simple powers tend to be extremely correlated when $K$ or $L$ increase. This does not affect the fit of the non-parametric procedure, but makes inferences on individual coefficients quite difficult. For this reason I will use Hermite polynomials, instead of Power series, as in Hausman et al. (1991). Normalized Hermite polynomials form an orthonormal basis in $L^2(\varphi)$, the Lebesgue space of $\mathbb{R}$-valued functions for which $f^2$ is $\varphi$ integrable. They are defined as:

$$H_n(x) = \frac{(-1)^n}{\sqrt{n!}} e^{\frac{x^2}{2}} \frac{d^n}{dx^n} \left( e^{-\frac{x^2}{2}} \right), \ n \geq 0$$

and obey the following recursive pattern:

$$H_0 = 1; \quad H_1 = x; \quad \sqrt{n+1} H_{n+1} = x H_n - \sqrt{n} H_{n-1}, \quad n \geq 2$$

Normalized Hermite polynomials also obey the following addition formula.

---

$^{19}$\(\varphi\) denotes the Gaussian measure.
\[ H_n(x + y) = \sum_{i=0}^{n} \frac{1}{2^i} \binom{n}{i}^{\frac{1}{2}} H_i\left(x\sqrt{2}\right) H_{n-i}\left(y\sqrt{2}\right) \] (3.33)

Lastly, it is easy to check that \( H_n \) as a coefficient \( 1/\sqrt{n!} \) on the term of order \( n \). In what follows, we adopt normalized Hermite polynomials, hence: \( p_i(.) = H_i(.) \).

**Estimating the reduced form parameters for a given \( K \) and \( L \).**

Assume for the time being that \( K \) and \( L \) are given. The method follows closely Hausman et al. (1991), adjusted for normalized Hermite polynomials. Substituting (3.32) in (3.23),

\[ k_i = \sum_{j=0}^{L} p_j(z_i) \alpha_j + v_i + \eta_i = p^L(z_i)' \alpha + v_i + \eta_i \]

where \( \alpha = (\alpha_0, ..., \alpha_L)' \) and \( p^L(z) = (H_0(z), H_1(z), ..., H_L(z))' \). Under (3.24) \( \alpha \) can be estimated by Ordinary Least Squares:

\[ \hat{\alpha} = \left(p^L(z)' p^L(z)\right)^{-1} p^L(z)' k \]

where, \( p^L_i = p^L(z_i) \) is a \((L + 1) \times 1\) vectors, \( z = (z_1, ..., z_n)' \), \( k = (k_1, ..., k_n)' \) are \( nx1 \) vectors and \( P^L(z) = (p^L_1, ..., p^L_n)' \) is a \( nx(L + 1) \) matrix. Define \( \hat{k}_i \) the fitted fundamental:

\[ \hat{k}(z_i) = p^L(z_i)' \hat{\alpha} \]

Since \( \alpha \) is identified, I follow Hausman et al. (1991) and assume that it is known in what follows.\(^{20}\)

Using the addition and recursion formulae for Hermite polynomials and following the

\(^{20}\)However, if \( \alpha \) is estimated, the covariance matrix of the second step will have to be adjusted for first step estimation. See Newey and MacFadden (1994).
derivation of (3.30) and (3.31), I show in the appendix that:

\[ e_t = p^K \left( \hat{k}_t \sqrt{2} \right)' \gamma + \mu_t \]  
\[ e_t k_t = p^{K+1} \left( \hat{k}_t \sqrt{2} \right)' \delta + \omega_t \]  
\[ E \left[ \mu_t | \hat{k}_t \right] = E \left[ \omega_t | \hat{k}_t \right] = 0 \]  

\( \gamma \) and \( \delta \) are a \( K+1 \) and \( K+2 \) vectors respectively. (3.36) implies that the reduced parameters \( \gamma \) and \( \delta \) can be estimated by Ordinary Least Squares. Of course, \( \gamma \) and \( \delta \) are linear combinations of the structural parameters \( \beta \) and the moments of the error term \( \nu_t \). These two equations are formally analog to (3.30) and (3.31) respectively. (3.35) is the auxiliary regression.

I also show in the appendix that the reduced parameters are defined by the following formulae:

for \( 0 \leq l \leq K \) :

\[ \gamma_l = \sum_{j=l}^{K} 2^{-\frac{l}{2}} \left( \begin{array}{l} j \\ l \end{array} \right)^{\frac{1}{2}} \hat{\rho}_{j-l} \beta_j \]  

(3.37)

for \( 1 \leq l \leq K - 1 \):

\[ \delta_l = \sum_{j=l}^{K-1} 2^{-\frac{j}{2}} \left( \sqrt{j} \beta_{j-1} + \sqrt{j + 1} \beta_{j+1} \right) \hat{\rho}_{j-l} \left( \begin{array}{l} j \\ l \end{array} \right)^{\frac{1}{2}} \]  
\[ + 2^{-\frac{K}{2}} \sqrt{K} \beta_{K-1} \hat{\rho}_{K-l} \left( \begin{array}{l} K \\ l \end{array} \right)^{\frac{1}{2}} \]  
\[ + 2^{-\frac{K+1}{2}} \sqrt{K+1} \beta_K \hat{\rho}_{K+1-l} \left( \begin{array}{l} K+1 \\ l \end{array} \right)^{\frac{1}{2}} \]  

(3.38)
\[
\delta_K = 2^{-\frac{K}{2}} \sqrt{K} \beta_{K-1} \\
\delta_{K+1} = 2^{-\frac{K+1}{2}} \sqrt{K+1} \beta_K
\]

where the \( \hat{\rho}_l \) are nuisance parameters linearly related to the moments of the error \( \nu_j = E\left[\nu_j^2\right] \), and defined in the appendix.\(^{21}\) It is easy to check that \( \hat{\rho}_0 = 1 \) and \( \hat{\rho}_1 = 0 \).

**Estimation of the structural and nuisance parameters for a given K and L;**

This implies that there are only \( 2K + 1 \) structural parameters to estimate: the \( K + 1 \) elements of \( \beta \) and the \( K \) remaining nuisance parameters \( (\hat{\rho}_2, ..., \hat{\rho}_{K+1}) \). We have \( 2K + 3 \) reduced parameters. Thus our procedure will generally be overidentified. However, I show in the appendix, following Hausman et al. (1991), that:

\[
\begin{align*}
\delta_{K+1} &= 2^{-\frac{K+1}{2}} \sqrt{K+1} \beta_K = \gamma_K \sqrt{\frac{K+1}{2}} ; \\
\delta_K &= 2^{-\frac{K}{2}} \sqrt{K} \beta_{K-1} = \gamma_{K-1} \sqrt{\frac{K}{2}} \\
\end{align*}
\]

(3.39)

Thus the last two coefficients of each reduced form vector form a separate block that over-identifies the higher order elements of \( \beta \). In the linear case, of course, this implies that all the coefficients are identified from the Instrumental Variable regression.

The general procedure for solving the inversion problem subject to the two overidentifying restrictions is presented in Hausman et al. (1991), once \( (\hat{\gamma}, \hat{\delta}) \) has been estimated.

First, \( \hat{\beta}_K \) and \( \hat{\beta}_{K-1} \) are estimated by a minimum chi-square. Then, the vector of remaining coefficients is exactly identified. More generally, if we write \( \pi' = (\gamma', \delta') \) the vector of reduced form coefficients, \( \Omega \) its variance covariance matrix and \( \pi = h \left( \begin{array}{c} \beta \\ \hat{\rho} \end{array} \right) \), where \( h(.) \) is defined by (3.37) and (3.38), then the problem can be restated as:

\(^{21}\)Specifically, the nuisance parameter \( \hat{\rho}_l \) is equal to \( E\left[\rho_l \left( \nu, \sqrt{2} \right)\right] \). Thus, it is a linear combination of the moments \( \nu \) up to order \( l \).
\[
\left( \begin{array}{c}
\hat{\beta} \\
\hat{\rho}
\end{array} \right) = \arg\min_{\hat{\beta}, \hat{\rho}} \left[ \pi - h \left( \begin{array}{c}
\hat{\beta} \\
\hat{\rho}
\end{array} \right) \right]' \Omega^{-1} \left[ \pi - h \left( \begin{array}{c}
\hat{\beta} \\
\hat{\rho}
\end{array} \right) \right]
\]

s.t. (3.39)

Lastly, if we define \( H = \partial h/\partial (\beta', \rho') \), \( \hat{H} \) its evaluation at \( \left( \begin{array}{c}
\hat{\beta} \\
\hat{\rho}
\end{array} \right) \) and \( \hat{\Omega} \) an estimate of the variance covariance of the reduced form parameters, the variance of the structural and nuisance parameters is estimated by:

\[
\text{var} \left( \begin{array}{c}
\beta \\
\rho
\end{array} \right) = \hat{H}^{-1} \hat{\Omega} \hat{H}'^{-1}
\]

This completes the description of the two-step procedure for a fixed \( K \) and \( L \).

Choosing optimally the order of approximation:

Even though this procedure is not efficient, we consider separately the choice of the instrument procedure (\( L \)) and of the main procedure (\( K \)).

Choosing \( L \): We can apply here all the standard Series Estimation results. In particular, if we define the Cross-Validation Criterion:\(^{22}\)

\[
CV = \sum \left( \frac{\hat{\nu}_t}{\hat{T}_t} \right)^2
\]

where:

\[
\nu_t = k_t - p^L(z_t) \alpha
\]

\(^{22}\)See Newey (1990)
\[ I_t = 1 - p^L(z_t)(P^L(z)'(P^L(z))^{-1}p^L(z_t)'. \]

then, under suitable regularity conditions, the rule "Choose L to minimize CV" is asymptotically optimal.\(^{23}\)

**Choosing K** The problem of choosing K is more complex. First, we have two regressions at our disposal, and probably want to use information from both of them. However, given the nuisance parameters, a simple Cross Validation criterion is unlikely to be asymptotically optimal. Intuitively, the bias due to the nuisance parameters is likely to dominate the variance component, given the error in variable problem. In empirical applications this conjecture seems to be verified as the CV criterion invariably returns a value \( K = 1 \). If this conjecture is true, one needs to increase the order of the approximation in order to reduce the bias. Since no asymptotic results are available yet for this situation, one has to rely on economic intuition. Without any additional results, I will adopt a conservative value of:

\[ K = 3 \]

This will still allow for non-linear effects up to a quadratic term to be captured by the estimation procedure. In addition, an approximation of at least cubic order is needed if one wants to allow for the S-shape implied by target-zone theory. In the future, I plan to investigate this issue more thoroughly using Monte Carlo simulations.

\(^{23}\)See Härdle and Linton (1993) and Newey (1990)
4 Empirical Results;

I now turn to the empirical implementation of the methodology outlined above. I first describe the data and the instrument. Then I report results from direct estimation, as in Flood et al. (1991). Lastly, I present the results from the NPIV procedure.

4.1 The data.

Andrew Rose provided me kindly with a complete copy of the data set used in Flood et al. (1991). It includes daily exchange rates and interest rates against the dollar for eight European countries (Germany, United Kingdom, France, Belgium, Netherlands, Denmark, Italy and Ireland) and for Japan, from March 1979 (the date of the creation of the Exchange Rate Mechanism) until May 1990.\textsuperscript{24} All these data come from the Bank of International Settlement. The exchange rates are quoted at the time of the official fixing against the dollar.\textsuperscript{25} For Exchange Rate Mechanism countries (ERM thereafter) we also have data on the central parities the ceilings and the floors. The interest rates are annualized “spot-to-next” rates, i.e. forward interest rates that apply from day 2 to day 3. Ideally, one would use overnight interest rates but these were not available at this frequency. The interest rates are simple bid rates at 10:00 am, Swiss time, averaged across several Euro markets, and thus free of political risk premia. All these data have been checked for errors and outliers by Flood et al. (1991). Each daily observation is treated identically, and no account is taken of time deformation as “day of the week” effects.

The sample can be divided into 13 regimes, within which no realignments occurred in the ERM. The dates of realignments are reported in Table 3.1.

\textsuperscript{24}Danish and Irish interest rates are not available before 1982.

\textsuperscript{25}Andrew Rose informed me that this fixing does not occur at the same time of the day for EMS countries, and that it would be preferable to use ecu-rates, which are smoother and quoted simultaneously. Unfortunately, I do not dispose of the latter.
Table 3.1: EMS REGIMES

<table>
<thead>
<tr>
<th>Regime</th>
<th>Starts</th>
<th># Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3-13-79</td>
<td>135</td>
</tr>
<tr>
<td>2</td>
<td>9-24-79</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>11-29-79</td>
<td>332</td>
</tr>
<tr>
<td>4</td>
<td>3-22-81</td>
<td>131</td>
</tr>
<tr>
<td>5</td>
<td>10-4-81</td>
<td>91</td>
</tr>
<tr>
<td>6</td>
<td>2-21-82</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>6-12-82</td>
<td>191</td>
</tr>
<tr>
<td>8</td>
<td>3-21-83</td>
<td>601</td>
</tr>
<tr>
<td>9</td>
<td>7-21-85</td>
<td>176</td>
</tr>
<tr>
<td>10</td>
<td>4-6-86</td>
<td>76</td>
</tr>
<tr>
<td>11</td>
<td>8-3-86</td>
<td>106</td>
</tr>
<tr>
<td>12</td>
<td>1-12-87</td>
<td>771</td>
</tr>
<tr>
<td>13</td>
<td>1-5-90</td>
<td>88</td>
</tr>
</tbody>
</table>

We can divide countries into two categories. The first one consists of countries belonging to the ERM: Denmark, Belgium, France, Netherlands, Italy and Ireland. Their exchange rate was allowed to fluctuate ±2.25% around the central parity.\textsuperscript{26} For these countries, we construct the bilateral exchange rate against the D-Mark. This seems reasonable, given the strong asymmetry in the ERM between Germany and other member countries. Also, as explained in section 2, our approach is very sensitive to the number of exogenous variables. The second category regroups floating regimes countries: Japan, UK and Germany.\textsuperscript{27} For these countries, we compute the bilateral exchange rate against the Dollar - which explains why Germany is a member of this category.

4.2 The Choice of the Instrument:

In order to implement the estimation procedure described in section 3, one needs as good an instrument as possible. At the frequency considered, there are not so many candi-

\textsuperscript{26} 5\% for Italy and Ireland in the beginning of the sample.

\textsuperscript{27} UK was a member of the European Monetary System, not of the ERM. The Sterling Pound was a component of the ECU basket, but the currency was allowed to float. Participation to the latter implies the obligation to defend official bands ±2.25\% around the central parity.
dates. More fundamentally, one can question whether any fundamental constructed from macroeconomic variables can capture the characteristics of the exchange rate process. Flood and Rose (1993), pursuing this line, argue that potential fundamental determinants should exhibit the same increase in conditional volatility as exchange rates, when moving from a fixed exchange rate regime to a free float. Empirically, there is no such macroeconomic variable.

In order to identify a correct instrument, one has to specify the error in variable. Recall from (2.12) that the constructed fundamental is defined as:

\[ k_t = e_t - \alpha (i_t - i_t^*) = k_t^* - \alpha \left( \frac{E[de_t]}{dt} - i_t + i_t^* \right) = k_t^* - \alpha R P_t \]

so that the error in variable is exactly proportional to the exchange risk premium \( R P_t \). Consequently, the instrument must be orthogonal to the exchange risk premium. At a daily frequency, one can think of using stock market returns differential as an instrument. Unfortunately, the stock return differential is likely to be correlated with the currency risk premium: assume that the risk premium increases, holding the expected depreciation constant. This implies that domestic interest rates have to rise with respect to foreign ones. Thus, by arbitrage, the return on stock at home is likely to rise with respect to the foreign stock market return.

However, we can consider the stock excess returns differential. Define the stock price in the domestic country at time \( t \) as \( p_t \). The annualized daily compounded return is then

\[ \lambda_t = 360 \ln \left( \frac{p_t}{p_{t-1}} \right) \]

The excess return differential is then defined as:

\[ R_t = (\lambda_t - i_t) - (\lambda_t^* - i_t^*) \]  \hspace{1cm} (4.40)
Table 3.2: Stock Market Indices

<table>
<thead>
<tr>
<th>Country</th>
<th>Index</th>
<th>Starts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>Copenhagen General</td>
<td>03/13/79</td>
</tr>
<tr>
<td>Belgium</td>
<td>Brussels General</td>
<td>01/02/80</td>
</tr>
<tr>
<td>France</td>
<td>CAC General</td>
<td>11/01/79</td>
</tr>
<tr>
<td>Netherlands</td>
<td>CBS All Shares</td>
<td>03/13/79</td>
</tr>
<tr>
<td>Italy</td>
<td>Milan Banca Comm</td>
<td>03/13/79</td>
</tr>
<tr>
<td>UK</td>
<td>FTSE</td>
<td>03/13/79</td>
</tr>
<tr>
<td>Japan</td>
<td>Nikkei</td>
<td>03/13/79</td>
</tr>
<tr>
<td>Germany</td>
<td>FAZ</td>
<td>03/13/79</td>
</tr>
<tr>
<td>US</td>
<td>S&amp;P Composite</td>
<td>03/13/79</td>
</tr>
</tbody>
</table>

where $i_t$ and $i_t^*$ are annualized one day interest rates. $(\lambda_t^* - i_t^*)$ represents the excess return on an investment in the foreign market, edged against currency risk. Thus, it is not surprising that it is unrelated to currency movements. In practice, the excess return differential, as defined in (4.40), appears to be extremely volatile and uncorrelated with the level of the exchange rate. We define the cumulated value according to:

$$z_t = \exp \left( \sum_{0}^{t} R_t / 360 \right)$$

Figure 4a-h plots this variable for all countries in our sample. Given the monetary interpretation given to the "true" fundamental, the differential excess stock return should be correlated with $k_t^*$. 

To construct the instrument variable, I gathered daily stock market indices for the all countries considered except Ireland.

---

28 One word of caution: in a general equilibrium setup, all returns on all assets are likely to be correlated. Therefore the currency risk premium, representing the return on a speculative investment in the foreign currency is likely to be correlated with the stock excess returns. Stricto sensu, our instrument is not valid in this case. However, we believe that there is so little evidence of correlation between stocks excess returns and currency risk that we can safely ignore this argument.

29 For some series, the initial frequency was weekly instead of daily. This led me to delete the first 212 observations of CBS, 16 observations of CAC, 12 observations of S&P, 112 observations of FAZ, 214 observations of Nikkei. I wish to thank Didier Valet from the Bank Indosuez in PARIS for providing me with these datas.
Table 3.3: DIRECT ESTIMATION: DENMARK

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
<th>$A_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.1933** (0.0852)</td>
<td>0.1940** (0.0857)</td>
<td>0.3247** (0.0009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-4.6963** (1.2779)</td>
<td>-4.7218** (1.3017)</td>
<td>-0.7965** (0.0533)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.0710 (0.9183)</td>
<td>-0.0500 (0.9611)</td>
<td>0.0965 (0.0725)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.6539** (0.2443)</td>
<td>1.6959** (0.2411)</td>
<td>0.2497** (0.0143)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-0.0680 (0.1553)</td>
<td>-0.0214 (0.1528)</td>
<td>-0.0998** (0.0064)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.1332 (0.2624)</td>
<td>0.1193 (0.2605)</td>
<td>-0.3976** (0.0103)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-0.7672** (0.0928)</td>
<td>-0.7361** (0.0908)</td>
<td>0.2403** (0.0046)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-0.3840** (0.1170)</td>
<td>-0.2511** (0.1162)</td>
<td>0.1564** (0.0075)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-0.3421** (0.1534)</td>
<td>-0.3705** (0.1542)</td>
<td>0.5948** (0.0062)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.

4.3 Direct Estimation;

This subsection directly estimate the non-linear terms, as in Flood et al. (1991):³⁰

$$e_t - k_t - \alpha \hat{\eta} = A_0 + A_1 \exp(\hat{\lambda}_1 k_t) + A_2 \exp(\hat{\lambda}_2 k_t) + A_3 k_t + u_t$$

I estimate (2.14) for each EMS regime independently, since the regimes may have had substantially different characteristics in terms of credibility and stability. Some small sample bias may arise as a result, since some regimes contain few observations (see Table 3.1). The results for $A_1, A_2$ and $A_3$ and their standard errors are reported in Tables 3.3-3.9. Given the strong serial correlation present in the data, I use Newey-West estimates for the variance covariance matrix, with 6 lags.

It appears that $A_1$ and $A_2$ are often significantly different from 0, both for EMS and floating countries. However the pattern of estimated coefficients exhibits an interesting feature: $A_1$ and $A_2$ are often similar in magnitude and of the same sign. These results indicate that honeymoon effects are present in the data at most at one of the boundary.

³⁰We estimate (2.14) without a constant, to avoid the quasi multicollinearity problem. Given that the constant was introduced to capture possible misspecifications, we can still estimate the true model under the null.
Table 3.4: DIRECT ESTIMATION: BELGIUM

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
<th>$A_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.8862**</td>
<td>(0.4621)</td>
<td>-1.9026**</td>
<td>(0.4608)</td>
<td>-0.0583**</td>
<td>(0.0115)</td>
</tr>
<tr>
<td>2</td>
<td>0.3859</td>
<td>(0.2576)</td>
<td>0.4243</td>
<td>(0.2538)</td>
<td>0.0937**</td>
<td>(0.0038)</td>
</tr>
<tr>
<td>3</td>
<td>-0.4752</td>
<td>(0.4822)</td>
<td>-0.3974</td>
<td>(0.4825)</td>
<td>0.1265**</td>
<td>(0.0112)</td>
</tr>
<tr>
<td>4</td>
<td>-1.9009*</td>
<td>(0.9526)</td>
<td>-1.7337*</td>
<td>(0.9616)</td>
<td>0.1056**</td>
<td>(0.0078)</td>
</tr>
<tr>
<td>5</td>
<td>-1.2503</td>
<td>(0.6925)</td>
<td>-1.2689</td>
<td>(0.6895)</td>
<td>-0.0286**</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>6</td>
<td>0.1433</td>
<td>(0.7407)</td>
<td>0.0839</td>
<td>(0.7347)</td>
<td>-0.5592**</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>7</td>
<td>0.6430**</td>
<td>(0.1343)</td>
<td>0.6090**</td>
<td>(0.1340)</td>
<td>0.0665**</td>
<td>(0.0036)</td>
</tr>
<tr>
<td>8</td>
<td>0.1396</td>
<td>(0.1492)</td>
<td>0.1699</td>
<td>(0.1475)</td>
<td>0.0908**</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>9</td>
<td>0.3013</td>
<td>(0.1946)</td>
<td>0.2706</td>
<td>(0.1956)</td>
<td>0.0395**</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>10</td>
<td>0.4900**</td>
<td>(0.1608)</td>
<td>0.5579**</td>
<td>(0.1632)</td>
<td>-0.3716**</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>11</td>
<td>-0.5321**</td>
<td>(0.1243)</td>
<td>-0.5164**</td>
<td>(0.1204)</td>
<td>0.0055</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>12</td>
<td>-0.1814*</td>
<td>(0.0997)</td>
<td>-0.1192</td>
<td>(0.0997)</td>
<td>0.0361**</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>13</td>
<td>-0.0995</td>
<td>(0.0684)</td>
<td>-0.1008</td>
<td>(0.0672)</td>
<td>0.2600**</td>
<td>(0.0015)</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.

Table 3.5: DIRECT ESTIMATION: FRANCE

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
<th>$A_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3301</td>
<td>(0.4086)</td>
<td>0.2721</td>
<td>(0.4215)</td>
<td>0.0621**</td>
<td>(0.0177)</td>
</tr>
<tr>
<td>2</td>
<td>0.2341</td>
<td>(0.2477)</td>
<td>0.2536</td>
<td>(0.2472)</td>
<td>0.7076**</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>3</td>
<td>-0.7828*</td>
<td>(0.3816)</td>
<td>-0.7954*</td>
<td>(0.3786)</td>
<td>0.4244**</td>
<td>(0.0203)</td>
</tr>
<tr>
<td>4</td>
<td>9.4816**</td>
<td>(1.4296)</td>
<td>9.8602**</td>
<td>(1.4069)</td>
<td>1.0683**</td>
<td>(0.1048)</td>
</tr>
<tr>
<td>5</td>
<td>0.4400**</td>
<td>(0.1738)</td>
<td>0.4589**</td>
<td>(0.1772)</td>
<td>0.0030</td>
<td>(0.0074)</td>
</tr>
<tr>
<td>6</td>
<td>17.4496**</td>
<td>(6.5263)</td>
<td>17.9680**</td>
<td>(6.4480)</td>
<td>1.1529**</td>
<td>(0.1471)</td>
</tr>
<tr>
<td>7</td>
<td>7.8050</td>
<td>(4.2806)</td>
<td>10.1350**</td>
<td>(4.1716)</td>
<td>6.0540**</td>
<td>(0.1108)</td>
</tr>
<tr>
<td>8</td>
<td>0.6210**</td>
<td>(0.1204)</td>
<td>0.6865**</td>
<td>(0.1199)</td>
<td>0.4598**</td>
<td>(0.0067)</td>
</tr>
<tr>
<td>9</td>
<td>-0.1658</td>
<td>(0.3938)</td>
<td>-0.1325</td>
<td>(0.3986)</td>
<td>0.3002**</td>
<td>(0.0122)</td>
</tr>
<tr>
<td>10</td>
<td>0.1349**</td>
<td>(0.0493)</td>
<td>0.1485**</td>
<td>(0.0503)</td>
<td>-0.7317**</td>
<td>(0.0023)</td>
</tr>
<tr>
<td>11</td>
<td>-0.1119</td>
<td>(0.1336)</td>
<td>-0.0949</td>
<td>(0.1398)</td>
<td>0.0983**</td>
<td>(0.0089)</td>
</tr>
<tr>
<td>12</td>
<td>-0.0189</td>
<td>(0.1163)</td>
<td>0.0289</td>
<td>(0.1172)</td>
<td>0.2719**</td>
<td>(0.0066)</td>
</tr>
<tr>
<td>13</td>
<td>0.2258</td>
<td>(0.3629)</td>
<td>0.2043</td>
<td>(0.3617)</td>
<td>0.6046**</td>
<td>(0.0062)</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.
### Table 3.6: Direct Estimation: Netherlands

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.1571)</td>
<td>(0.1248)</td>
</tr>
<tr>
<td>1</td>
<td>0.2803</td>
<td>(0.1620)</td>
<td>0.2401</td>
</tr>
<tr>
<td>2</td>
<td>0.5101</td>
<td>(0.3236)</td>
<td>0.4952</td>
</tr>
<tr>
<td>3</td>
<td>-1.6775*(0.5130)</td>
<td>-1.7016**(0.5111)</td>
<td>-2.9921**(0.2454)</td>
</tr>
<tr>
<td>4</td>
<td>0.5182</td>
<td>(0.5463)</td>
<td>0.5629</td>
</tr>
<tr>
<td>5</td>
<td>1.5164**(0.4676)</td>
<td>1.5059**(0.4701)</td>
<td>6.8272**(0.1306)</td>
</tr>
<tr>
<td>6</td>
<td>-3.5902** (1.0921)</td>
<td>-3.5467** (1.0730)</td>
<td>-4.7673** (0.3742)</td>
</tr>
<tr>
<td>7</td>
<td>-0.4478**(0.2019)</td>
<td>-0.4424** (0.1978)</td>
<td>-1.8406** (0.0939)</td>
</tr>
<tr>
<td>8</td>
<td>0.0874</td>
<td>(0.1034)</td>
<td>0.0817</td>
</tr>
<tr>
<td>9</td>
<td>-0.0015</td>
<td>(0.1074)</td>
<td>0.0104</td>
</tr>
<tr>
<td>10</td>
<td>-0.1876</td>
<td>(0.1852)</td>
<td>-0.1453</td>
</tr>
<tr>
<td>11</td>
<td>0.0983</td>
<td>(0.1532)</td>
<td>0.1035</td>
</tr>
<tr>
<td>12</td>
<td>-0.7048** (0.1833)</td>
<td>-0.6786** (0.1812)</td>
<td>-1.6655** (0.0610)</td>
</tr>
<tr>
<td>13</td>
<td>-0.1634</td>
<td>(0.0979)</td>
<td>-0.1638</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.

### Table 3.7: Direct Estimation: Italy

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(0.9718)</td>
<td>(0.0055)</td>
</tr>
<tr>
<td>1</td>
<td>1.8261**(0.9489)</td>
<td>1.8750**(1.0586)</td>
<td>0.1630** (0.0053)</td>
</tr>
<tr>
<td>2</td>
<td>1.4329</td>
<td>(1.0606)</td>
<td>1.5222</td>
</tr>
<tr>
<td>3</td>
<td>3.2214</td>
<td>(1.9189)</td>
<td>3.3685</td>
</tr>
<tr>
<td>4</td>
<td>10.1855**(1.7384)</td>
<td>11.1587**(1.6958)</td>
<td>0.1252** (0.0109)</td>
</tr>
<tr>
<td>5</td>
<td>-1.6309** (0.6989)</td>
<td>-1.6067** (1.9578)</td>
<td>0.1361** (0.0056)</td>
</tr>
<tr>
<td>6</td>
<td>3.0335</td>
<td>(4.9381)</td>
<td>3.1671</td>
</tr>
<tr>
<td>7</td>
<td>18.4395**(1.1400)</td>
<td>19.3383**(1.0614)</td>
<td>0.2686** (0.0114)</td>
</tr>
<tr>
<td>8</td>
<td>-0.0878**(0.0334)</td>
<td>-0.0517 (0.0370)</td>
<td>0.0539** (0.0012)</td>
</tr>
<tr>
<td>9</td>
<td>0.1268**(0.4260)</td>
<td>0.0391</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.8153**(0.1721)</td>
<td>0.9554** (0.4256)</td>
<td>0.0832** (0.0052)</td>
</tr>
<tr>
<td>11</td>
<td>0.4729</td>
<td>(0.3952)</td>
<td>0.5692</td>
</tr>
<tr>
<td>12</td>
<td>-0.0453</td>
<td>(0.0987)</td>
<td>-0.0517</td>
</tr>
<tr>
<td>13</td>
<td>0.5935</td>
<td>(0.7616)</td>
<td>0.5345</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.
Table 3.8: DIRECT ESTIMATION: IRELAND

<table>
<thead>
<tr>
<th>Regime</th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
<th>$A_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-2.0381**</td>
<td>(0.5658)</td>
<td>-2.0499*</td>
<td>(0.5617)</td>
<td>0.2824**</td>
<td>(0.0108)</td>
</tr>
<tr>
<td>7</td>
<td>1.5935**</td>
<td>(0.2906)</td>
<td>1.5667**</td>
<td>(0.2916)</td>
<td>-0.2042**</td>
<td>(0.0134)</td>
</tr>
<tr>
<td>8</td>
<td>0.6058**</td>
<td>(0.1731)</td>
<td>0.7200**</td>
<td>(0.1793)</td>
<td>-0.5696**</td>
<td>(0.0087)</td>
</tr>
<tr>
<td>9</td>
<td>2.1608</td>
<td>(0.3114)</td>
<td>2.0121**</td>
<td>(0.3037)</td>
<td>-0.0105</td>
<td>(0.0182)</td>
</tr>
<tr>
<td>10</td>
<td>0.8372**</td>
<td>(0.2771)</td>
<td>0.9126**</td>
<td>(0.2823)</td>
<td>0.9832**</td>
<td>(0.0165)</td>
</tr>
<tr>
<td>11</td>
<td>0.2749</td>
<td>(0.4857)</td>
<td>0.1816</td>
<td>(0.4804)</td>
<td>0.0913**</td>
<td>(0.0291)</td>
</tr>
<tr>
<td>12</td>
<td>0.9880**</td>
<td>(0.2233)</td>
<td>1.1146**</td>
<td>(0.2374)</td>
<td>-0.4430**</td>
<td>(0.0199)</td>
</tr>
<tr>
<td>13</td>
<td>-0.0778</td>
<td>(0.5240)</td>
<td>-0.030i</td>
<td>(0.5243)</td>
<td>-1.1179**</td>
<td>(0.0092)</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.

Table 3.9: DIRECT ESTIMATION: NON EMS COUNTRIES

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
<th>$A_3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>0.0800**</td>
<td>(0.0398)</td>
<td>0.0691**</td>
<td>(0.0471)</td>
<td>0.1548**</td>
<td>(0.0296)</td>
</tr>
<tr>
<td>Japan</td>
<td>-0.2633**</td>
<td>(0.0275)</td>
<td>-0.2289**</td>
<td>(0.0254)</td>
<td>0.0343**</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>Germany</td>
<td>0.1690**</td>
<td>(0.0288)</td>
<td>0.1367**</td>
<td>(0.0295)</td>
<td>-0.3349**</td>
<td>(0.0124)</td>
</tr>
</tbody>
</table>

All variables and standard errors are multiplied by 100. Robust standard errors.
Table 3.10: T-Test of $H_0$: $A_1A_2 > 0$, EMS Countries

<table>
<thead>
<tr>
<th>Regime</th>
<th>Denmark</th>
<th>Belgium</th>
<th>France</th>
<th>Netherlands</th>
<th>Italy</th>
<th>Ireland</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.</td>
<td>2.0531</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>0.7902</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>0.4489</td>
<td>1.0384</td>
<td>1.6500</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>.</td>
<td>0.9472</td>
<td>3.4080</td>
<td>0.4940</td>
<td>3.0996</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>.</td>
<td>0.9114</td>
<td>1.2804</td>
<td>1.6115</td>
<td>1.1562</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>1.8258</td>
<td>0.0718</td>
<td>1.3645</td>
<td>1.6483</td>
<td>0.3100</td>
<td>1.8130</td>
</tr>
<tr>
<td>7</td>
<td>0.0311</td>
<td>2.3338</td>
<td>1.0416</td>
<td>1.1138</td>
<td>8.5692</td>
<td>2.7148</td>
</tr>
<tr>
<td>8</td>
<td>3.4506</td>
<td>0.5165</td>
<td>2.7143</td>
<td>0.4104</td>
<td>0.9159</td>
<td>1.8703</td>
</tr>
<tr>
<td>9</td>
<td>0.1062</td>
<td>0.7307</td>
<td>0.1858</td>
<td>-0.0158</td>
<td>0.0702</td>
<td>3.3910</td>
</tr>
<tr>
<td>10</td>
<td>0.2407</td>
<td>1.6113</td>
<td>1.4198</td>
<td>0.4427</td>
<td>2.6145</td>
<td>1.5620</td>
</tr>
<tr>
<td>11</td>
<td>4.0947</td>
<td>2.1428</td>
<td>0.3751</td>
<td>0.3295</td>
<td>0.6473</td>
<td>0.2267</td>
</tr>
<tr>
<td>12</td>
<td>1.3039</td>
<td>0.7219</td>
<td>-0.4748</td>
<td>1.8972</td>
<td>0.2473</td>
<td>2.2788</td>
</tr>
<tr>
<td>13</td>
<td>1.1571</td>
<td>0.7388</td>
<td>0.2960</td>
<td>0.8305</td>
<td>0.3684</td>
<td>0.0414</td>
</tr>
</tbody>
</table>

To confirm this intuition, I propose a test of the hypothesis $H_0$: $A_1A_2 > 0$. Using the delta method, and denoting $\beta$ the vector $(A_1A_2A_3)'$:

$$\hat{\beta} \sim_{as} N(\beta, V) \Rightarrow g(\hat{\beta}) \sim_{as} N(g(\beta), \partial g/\partial \beta' V \partial g/\partial \beta)$$ (4.41)

where $V$ is the asymptotic variance of $\beta$. I test the null hypothesis by using a t-test, asymptotically distributed as a normal. Results of this test are reported in table 3.10-3.11.

Typically, the hypothesis that $A_1$ and $A_2$ have the same sign is never rejected.

A formal test of the target zone theory would be a test of $H_0$: $A_1 < 0$ and $A_2 > 0$. However, joint one-sided tests are difficult to construct. Therefore I will confine myself to tests of $H_0$: $A_1 < 0$ and $H_0'$: $A_2 > 0$ each in turn. Under $H_0$, $A_1/\sigma_1$ follows a t-statistic. The critical value for a test with 5% confidence level is 1.66 for $H_0$ and $-1.66$ for $H_0'$.\textsuperscript{31}

$H_0$ tests for stabilizing effects in the upper part of the band, i.e. when a given

\textsuperscript{31}The information to carry this test is already present in Tables 3.3-3.9

148
Table 3.11: T-Test of $H0$: $A_1A_2 > 0$, Non-EMS Countries

<table>
<thead>
<tr>
<th>United Kingdom</th>
<th>Japan</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8614</td>
<td>4.7162</td>
<td>2.6203</td>
</tr>
</tbody>
</table>

EMS currency depreciates against the DM, while $H0'$ tests for stabilizing effects in the lower part, when a currency appreciates against the DM. The results confirm that there is some asymmetry: in 16 cases, we are able to find significant negative $A_1$, most notably for Denmark, Belgium and Netherlands. This could reflect the fact that these countries belong to the “core” EMS and have devalued relatively infrequently vis a vis the Deutschemark. In 20 cases, we find evidence of stabilizing effects at the lower boundary, most notably for France, Italy and Ireland. This result is more subtle to interpret. These countries devalued relatively frequently against the Deutschemark and almost never reached the lower boundary of their notional band. Given our earlier observation that $A_1$ and $A_2$ tend to be of the same sign, this result indicates rather that $A_1$ is positive for these countries. The target zone seem to have exerted a destabilizing effect. Indeed, looking directly at Tables 3.3-3.9, $A_1$ is positive and significant in 5 regimes out of 13 for France and Italy, 4 regimes out of 8 for Ireland, but only in 2 cases for Denmark and Belgium, and 1 case for the Netherlands. These results, taken together, seem to indicate a strong asymmetry in the ERM between a de facto DM area (Belgium, Denmark, Netherlands) and peripheral countries, whose currencies suffered repeated devaluations.

Results for floating countries are harder to rationalize. Although none of the currencies was allowed to float freely against the Dollar, strong non-linear effects are present, of about the same order of magnitude as the managed exchange rates.

4.4 Nonparametric Results

We apply in this section the methodology presented above. The first step consist in estimating $\alpha$, the vector of coefficients for the instrumental equation. The fitted variable
Table 3.12: Reduced Parameters, Denmark

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1.1604**</td>
<td>(0.038)</td>
<td>2.8054**</td>
<td>(0.685)</td>
<td>-0.1351**</td>
<td>(0.055)</td>
<td>2.2616**</td>
<td>(0.565)</td>
</tr>
<tr>
<td>8</td>
<td>1.4223**</td>
<td>(0.051)</td>
<td>-3.2140</td>
<td>(2.013)</td>
<td>0.1968**</td>
<td>(0.072)</td>
<td>-2.7286</td>
<td>(1.649)</td>
</tr>
<tr>
<td>9</td>
<td>1.3760**</td>
<td>(0.045)</td>
<td>-2.0740**</td>
<td>(0.632)</td>
<td>0.1173*</td>
<td>(0.064)</td>
<td>-1.7610**</td>
<td>(0.518)</td>
</tr>
<tr>
<td>11</td>
<td>1.3193**</td>
<td>(0.008)</td>
<td>0.4854**</td>
<td>(0.135)</td>
<td>-0.0133</td>
<td>(0.012)</td>
<td>0.3959**</td>
<td>(0.113)</td>
</tr>
<tr>
<td>12</td>
<td>1.0730**</td>
<td>(0.055)</td>
<td>-3.5075</td>
<td>(2.479)</td>
<td>-0.3876**</td>
<td>(0.079)</td>
<td>-3.0046</td>
<td>(2.030)</td>
</tr>
</tbody>
</table>

Table 3.13: Reduced Parameters, Belgium

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.7800**</td>
<td>(0.034)</td>
<td>1.7298**</td>
<td>(0.856)</td>
<td>0.0018</td>
<td>(0.049)</td>
<td>1.3781*</td>
<td>(0.704)</td>
</tr>
<tr>
<td>7</td>
<td>2.8488**</td>
<td>(0.045)</td>
<td>-0.9506</td>
<td>(1.008)</td>
<td>-0.1669**</td>
<td>(0.064)</td>
<td>-0.8623</td>
<td>(0.828)</td>
</tr>
<tr>
<td>8</td>
<td>2.8473**</td>
<td>(0.025)</td>
<td>-7.8560**</td>
<td>(0.877)</td>
<td>-0.2249**</td>
<td>(0.036)</td>
<td>-6.5263**</td>
<td>(0.719)</td>
</tr>
<tr>
<td>9</td>
<td>2.9922**</td>
<td>(0.011)</td>
<td>-1.7446**</td>
<td>(0.250)</td>
<td>-0.0292*</td>
<td>(0.016)</td>
<td>-1.4725**</td>
<td>(0.206)</td>
</tr>
<tr>
<td>11</td>
<td>3.0308**</td>
<td>(0.004)</td>
<td>-0.2181</td>
<td>(0.155)</td>
<td>-0.0023</td>
<td>(0.006)</td>
<td>-0.1915</td>
<td>(0.132)</td>
</tr>
<tr>
<td>12</td>
<td>2.5735**</td>
<td>(0.043)</td>
<td>-10.8447**</td>
<td>(1.679)</td>
<td>-0.6601**</td>
<td>(0.061)</td>
<td>-8.9202**</td>
<td>(1.373)</td>
</tr>
</tbody>
</table>

$\hat{k}$ is then used as a regressor in (3.30) and (3.31). Since, I impose $K = 3$, the main and auxiliary regression will directly give us estimates of $\beta_2$ and $\beta_3$, the coefficients on the Hermite polynomials of order 2 and 3, according to (3.39). In this first pass at the data, I will only report results for these two parameters.\(^{32}\)

The results come in a series of graphics and tables. Tables 3.12-3.17 report the estimates for the reduced parameters $\gamma$ in (3.34) and their standard errors.\(^{33}\) As discussed in the previous section, these parameters differ from the structural parameters $\beta$. However, it seems interesting to compare the two of them. The results are reported only for the regimes 3, 7, 8, 9, 11 and 12 for which the largest number of observations on both the exchange rate and the instrument were available.

\(^{32}\)Although the recursion (3.37) and (3.38) is easy to solve, the formulae for the correct variance-covariance of the structural parameters is substantially more involved.

\(^{33}\)Both corrected for first step estimation and robust to potential heteroscedasticity.
### Table 3.14: Reduced Parameters, France

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.0066**</td>
<td>0.024</td>
<td>-2.7206**</td>
<td>0.649</td>
<td>0.2320**</td>
<td>0.035</td>
<td>-2.2658**</td>
<td>0.534</td>
</tr>
<tr>
<td>7</td>
<td>0.6919**</td>
<td>0.034</td>
<td>2.3003**</td>
<td>0.670</td>
<td>-0.4925**</td>
<td>0.049</td>
<td>1.8547**</td>
<td>0.553</td>
</tr>
<tr>
<td>8</td>
<td>0.2999**</td>
<td>0.067</td>
<td>-17.2764**</td>
<td>1.345</td>
<td>-1.1558**</td>
<td>0.095</td>
<td>-14.2114**</td>
<td>1.100</td>
</tr>
<tr>
<td>9</td>
<td>1.1415**</td>
<td>0.016</td>
<td>-0.8824**</td>
<td>0.266</td>
<td>0.0335</td>
<td>0.023</td>
<td>-0.7485**</td>
<td>0.218</td>
</tr>
<tr>
<td>11</td>
<td>1.1815**</td>
<td>0.007</td>
<td>0.3025**</td>
<td>0.077</td>
<td>-0.0063</td>
<td>0.010</td>
<td>0.2454**</td>
<td>0.065</td>
</tr>
<tr>
<td>12</td>
<td>0.8745**</td>
<td>0.070</td>
<td>-8.8097**</td>
<td>3.238</td>
<td>-0.4852**</td>
<td>0.100</td>
<td>-7.3088**</td>
<td>2.648</td>
</tr>
</tbody>
</table>

### Table 3.15: Reduced Parameters, Netherlands

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.1975**</td>
<td>0.019</td>
<td>-4.8016**</td>
<td>0.519</td>
<td>0.1540**</td>
<td>0.028</td>
<td>-3.9992**</td>
<td>0.427</td>
</tr>
<tr>
<td>7</td>
<td>0.0587**</td>
<td>0.013</td>
<td>-1.7823**</td>
<td>0.395</td>
<td>-0.0512**</td>
<td>0.019</td>
<td>-1.5031**</td>
<td>0.325</td>
</tr>
<tr>
<td>8</td>
<td>-0.0652**</td>
<td>0.030</td>
<td>8.8760**</td>
<td>1.854</td>
<td>-0.2612**</td>
<td>0.043</td>
<td>7.2236**</td>
<td>1.157</td>
</tr>
<tr>
<td>9</td>
<td>0.1189**</td>
<td>0.006</td>
<td>-0.2486</td>
<td>0.191</td>
<td>-0.0009</td>
<td>0.008</td>
<td>-0.2158</td>
<td>0.158</td>
</tr>
<tr>
<td>11</td>
<td>0.1149**</td>
<td>0.003</td>
<td>0.0404</td>
<td>0.052</td>
<td>-0.0098*</td>
<td>0.005</td>
<td>0.0299</td>
<td>0.044</td>
</tr>
<tr>
<td>12</td>
<td>-0.0109</td>
<td>0.028</td>
<td>-3.4242**</td>
<td>0.795</td>
<td>-0.1847**</td>
<td>0.040</td>
<td>-2.8224**</td>
<td>0.650</td>
</tr>
</tbody>
</table>

### Table 3.16: Reduced Parameters, Italy

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5.9312**</td>
<td>0.052</td>
<td>-3.2607</td>
<td>2.128</td>
<td>-0.3218**</td>
<td>0.075</td>
<td>-2.7611</td>
<td>1.745</td>
</tr>
<tr>
<td>7</td>
<td>6.2321**</td>
<td>0.078</td>
<td>-1.1785</td>
<td>2.895</td>
<td>-0.1625</td>
<td>0.111</td>
<td>-1.0611</td>
<td>2.375</td>
</tr>
<tr>
<td>8</td>
<td>6.5771**</td>
<td>0.071</td>
<td>12.8730**</td>
<td>4.926</td>
<td>0.2235**</td>
<td>0.101</td>
<td>10.2870**</td>
<td>4.035</td>
</tr>
<tr>
<td>9</td>
<td>6.4745**</td>
<td>0.029</td>
<td>-1.9123**</td>
<td>0.475</td>
<td>-0.0599</td>
<td>0.042</td>
<td>-1.6198**</td>
<td>0.392</td>
</tr>
<tr>
<td>11</td>
<td>6.5285**</td>
<td>0.004</td>
<td>-0.0440</td>
<td>0.079</td>
<td>-0.0149**</td>
<td>0.006</td>
<td>-0.0461</td>
<td>0.066</td>
</tr>
<tr>
<td>12</td>
<td>6.2336**</td>
<td>0.063</td>
<td>-17.9901**</td>
<td>3.501</td>
<td>-0.5110**</td>
<td>0.089</td>
<td>-14.8907**</td>
<td>2.949</td>
</tr>
</tbody>
</table>

### Table 3.17: Reduced Parameters, Non-EMS Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>$\gamma_0$</th>
<th>S.E.</th>
<th>$\gamma_1$</th>
<th>S.E.</th>
<th>$\gamma_2$</th>
<th>S.E.</th>
<th>$\gamma_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>7.23</td>
<td>4.96</td>
<td>-1831.14**</td>
<td>523.88</td>
<td>10.85</td>
<td>7.03</td>
<td>-1498.33**</td>
<td>428.03</td>
</tr>
<tr>
<td>Germany</td>
<td>61.03</td>
<td>46.07</td>
<td>-3077.88</td>
<td>2362.60</td>
<td>79.06</td>
<td>65.25</td>
<td>-2521.55</td>
<td>1929.98</td>
</tr>
<tr>
<td>Japan</td>
<td>107.81</td>
<td>74.76</td>
<td>1570.71</td>
<td>2684.75</td>
<td>151.48</td>
<td>105.76</td>
<td>1277.19</td>
<td>2191.52</td>
</tr>
</tbody>
</table>

151
Table 3.18: Structural Parameters, Denmark

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-0.2647**</td>
<td>(0.1130)</td>
<td>6.5151**</td>
<td>(1.5945)</td>
</tr>
<tr>
<td>8</td>
<td>0.3917**</td>
<td>(0.1472)</td>
<td>-7.8799</td>
<td>(4.7729)</td>
</tr>
<tr>
<td>9</td>
<td>0.2420**</td>
<td>(0.1298)</td>
<td>-5.2244**</td>
<td>(1.4686)</td>
</tr>
<tr>
<td>11</td>
<td>-0.0223</td>
<td>(0.0242)</td>
<td>1.0988**</td>
<td>(0.3159)</td>
</tr>
<tr>
<td>12</td>
<td>-0.7898**</td>
<td>(0.1875)</td>
<td>-8.7770</td>
<td>(6.3334)</td>
</tr>
</tbody>
</table>

Table 3.19: Structural Parameters, Belgium

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.0035</td>
<td>(0.0992)</td>
<td>4.0268**</td>
<td>(2.0034)</td>
</tr>
<tr>
<td>7</td>
<td>-0.3371**</td>
<td>(0.1305)</td>
<td>-2.3743</td>
<td>(2.3776)</td>
</tr>
<tr>
<td>8</td>
<td>-0.4585**</td>
<td>(0.0764)</td>
<td>-18.5681**</td>
<td>(2.1807)</td>
</tr>
<tr>
<td>9</td>
<td>-0.0626*</td>
<td>(0.0334)</td>
<td>-4.1965**</td>
<td>(0.5973)</td>
</tr>
<tr>
<td>11</td>
<td>-0.0053</td>
<td>(0.0129)</td>
<td>-0.5588</td>
<td>(0.3788)</td>
</tr>
<tr>
<td>12</td>
<td>-1.3156**</td>
<td>(0.1254)</td>
<td>-24.7994**</td>
<td>(4.0124)</td>
</tr>
</tbody>
</table>

The results indicate that strong non-linearities are present. Figure 5.1–6 present the fitted exchange rate as a function of the instrumented fundamental $\hat{k}$. These figures indicate that the familiar S-shape is present for numerous countries and regimes, most notably Belgium, Netherlands and, to some extent, France. By contrast, the estimates for the floating countries are large and not significant, except for Britain. At face value, these results would indicate that target zone models do perform quite well. However, we need to assess those results on the structural parameters in order to validate this intuition. Table 3.18-3.23 report the estimates for the quadratic and cubic terms $\beta_2$ and $\beta_3$, together with their standard deviation.

The results indicate that the non-linearities are very substantial. This is not surprising: under the hypothesis that the model is well specified, $\gamma_2$, $\gamma_3$, $\beta_2$ and $\beta_3$ are related through (3.39). In particular, it is straightforward to verify that $\gamma_2$ and $\beta_2$ have the same sign and are simultaneously significant. The same holds for $\gamma_3$ and $\beta_3$. Obviously, these results hold since the only nuisance parameters to affect $\gamma_2$ and $\gamma_3$ are $\hat{\rho}_0$ and $\hat{\rho}_1$. 

152
Table 3.20: Structural Parameters, France

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.4579**</td>
<td>(0.0727)</td>
<td>-6.7682**</td>
<td>(1.5885)</td>
</tr>
<tr>
<td>7</td>
<td>-1.0060**</td>
<td>(0.0911)</td>
<td>5.7287**</td>
<td>(1.4633)</td>
</tr>
<tr>
<td>8</td>
<td>-2.3035**</td>
<td>(0.1991)</td>
<td>-39.2528**</td>
<td>(3.2636)</td>
</tr>
<tr>
<td>9</td>
<td>0.0625</td>
<td>(0.0464)</td>
<td>-2.1354**</td>
<td>(0.6208)</td>
</tr>
<tr>
<td>11</td>
<td>-0.0048</td>
<td>(0.0216)</td>
<td>0.7056**</td>
<td>(0.1827)</td>
</tr>
<tr>
<td>12</td>
<td>-0.9636**</td>
<td>(0.2077)</td>
<td>-19.1854**</td>
<td>(7.8085)</td>
</tr>
</tbody>
</table>

Table 3.21: Structural Parameters, Netherlands

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.1472**</td>
<td>(0.0570)</td>
<td>-10.8532**</td>
<td>(1.1503)</td>
</tr>
<tr>
<td>7</td>
<td>-0.1576**</td>
<td>(0.0358)</td>
<td>-2.1364**</td>
<td>(0.8393)</td>
</tr>
<tr>
<td>8</td>
<td>-0.4824**</td>
<td>(0.0874)</td>
<td>21.8588**</td>
<td>(4.3354)</td>
</tr>
<tr>
<td>9</td>
<td>-0.0087</td>
<td>(0.0171)</td>
<td>-0.2805</td>
<td>(0.4122)</td>
</tr>
<tr>
<td>11</td>
<td>-0.0137</td>
<td>(0.0091)</td>
<td>0.0347</td>
<td>(0.1151)</td>
</tr>
<tr>
<td>12</td>
<td>-0.3226**</td>
<td>(0.0795)</td>
<td>-5.4277**</td>
<td>(1.8042)</td>
</tr>
</tbody>
</table>

Table 3.22: Structural Parameters, Italy

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-0.6458**</td>
<td>(0.1554)</td>
<td>-7.6010</td>
<td>(4.8950)</td>
</tr>
<tr>
<td>7</td>
<td>-0.3287</td>
<td>(0.2224)</td>
<td>-2.9013</td>
<td>(6.8212)</td>
</tr>
<tr>
<td>8</td>
<td>0.4473**</td>
<td>(0.2109)</td>
<td>29.1592**</td>
<td>(11.0116)</td>
</tr>
<tr>
<td>9</td>
<td>-0.1183</td>
<td>(0.0878)</td>
<td>-4.5299**</td>
<td>(1.1504)</td>
</tr>
<tr>
<td>11</td>
<td>-0.0299**</td>
<td>(0.0122)</td>
<td>-0.1308</td>
<td>(0.1882)</td>
</tr>
<tr>
<td>12</td>
<td>-1.0247**</td>
<td>(0.2120)</td>
<td>-42.2699**</td>
<td>(8.8780)</td>
</tr>
</tbody>
</table>

Table 3.23: Structural Parameters, Non-EMS Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>$\beta_2$</th>
<th>S.E.</th>
<th>$\beta_3$</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>22.5089**</td>
<td>(4.6781)</td>
<td>-4154.5092**</td>
<td>(333.7805)</td>
</tr>
<tr>
<td>Germany</td>
<td>157.7716**</td>
<td>(11.4987)</td>
<td>-7144.9238**</td>
<td>(600.0612)</td>
</tr>
<tr>
<td>Japan</td>
<td>300.6335**</td>
<td>(17.6764)</td>
<td>3437.2371**</td>
<td>(647.8174)</td>
</tr>
</tbody>
</table>

153
whose values are already known. The estimated $\hat{\beta}_2$ and $\hat{\beta}_3$ differ from the corresponding reduced form estimated coefficients $\gamma$ since I use information from the auxiliary regression, $\delta$, to correct for possible misspecification. The similarity implies that the analysis of the results is pretty much unchanged. In particular, S-curve like non-linearities will be associated with a negative $\beta_3$. The results indicate that the exchange rates for EMS countries benefitted from stabilizing effects in 13 out of 29 cases. In 7 cases, the non-linear terms were outright destabilizing ($\beta_3 > 0$).

For EMS countries, therefore, very significant non-linear effects appear in the data, contrary to previous studies. These effects, as shown on Figure 5.1-6, exhibit for some regimes the very S-shape predicted by the basic theory. In particular, Netherlands seem to have benefited from stabilizing effects both at the upper and lower boundaries in almost every regime. This conforms to one's expectation, given the close link between the Florin and the Mark. This result is also in accordance with the direct estimation results presented in the previous subsection: the Netherland belongs the de facto to the “DM-currency area.” There is, however, a fundamental difference between NPIV and direct estimation results. As discussed previously, direct estimation results suggested that the stabilizing non-linearities were present at most at one edge of the band. This in turn, is in accordance with the argument that the EMS is an asymmetric monetary system.

NPIV results tell an entirely different story. For some countries and regimes (e.g. Denmark in regimes 7 and 11, Belgium in regime 3, France in regime 11, Netherlands and Italy in regime 8) the non-linearities are destabilizing at both end. In the majority of cases, however (Denmark in regimes 8, 9 an 12; Belgium in regimes 7, 8, 9 and 12; France in regimes 3, 8, 9 and 12; Netherlands throughout the period; and Italy in regimes 3, 7, 9, 11 and 12), the exchange rate is stabilized at least at one edge of the band, frequently at both, as in the standard theory. These results are strongly supporting.

The results for the floating countries are somewhat more mystifying. According to
the reduced parameters, there are almost no non-linear effects. However, the structural parameters are radically different. $\beta_7$ and $\beta_3$ are large and significant. I find the magnitude somewhat implausible (see Table 3.23). Since estimation for the floating countries is performed on the full sample, there is a good possibility that the approximation should be of higher order. Estimating the exchange rate process with higher order approximations requires the full resolution of the recursion algorithm. I plan to do such estimation in the future.

5 Conclusion

This chapter proposed a Non-Parametric Instrumental Variable Estimator to overcome the fundamental's unobservability in target-zone theory. Using the compounded stock market excess return differential as an instrument, this method gives consistent estimators of the non-linear terms in the exchange rate process. Preliminary results show that non-linearities are substantial and conform to a surprising extent to the predictions of the basic target zone theory. This result highlights the dangers of using mismeasured regressors in non-linear models. The approach developed is versatile enough to solve similar problems in a variety of situations. Investment dynamics, were the user cost is typically constructed from capital stock data, face the same type of problems. Employment dynamics, driven partially by were fluctuations in the opportunity cost of labor, would also benefit from the methodology developed in this chapter.
Appendix B

a  Derivations

a.1  Derivation of (3.34):

Define the sequence \( \{a_{ij}\} \) such that

\[
p_j(z) = \sum_{i=0}^{j} a_{ij} z^i
\]  \hspace{1cm} (a.1)

The \( a_{ij} \) are the coefficients of the projection of the Hermite polynomials on the Power series.

Let assume that \( \alpha \) is known in this derivation. We write

\[
e_t = \sum_{j=0}^{K} \beta_j p_j(w_t + v_t) + \epsilon_t
\]  \hspace{1cm} (a.2)

where \( w_t = \sum_{i=0}^{L} p_i(z_t) \alpha_i \). In practice, we do not observe the true \( \alpha \)'s. We will use \( \hat{w}_t = \sum_{i=0}^{L} p_i(z_t) \hat{\alpha}_i \), using the \( \alpha \)'s estimated in the first step. Using the addition formula for normalized Hermite polynomials, one can rewrite (a.2) as

\[
e_t = \sum_{j=0}^{K} \beta_j \sum_{i=0}^{j} 2^{-\frac{j}{2}} \binom{j}{i}^{1/2} p_i(w_t \sqrt{2}) p_{j-i}(v_t \sqrt{2}) + \epsilon_t
\]

\[
= \sum_{l=0}^{K} p_l(w_t \sqrt{2}) \sum_{j=l}^{K} \beta_j 2^{-\frac{j}{2}} \binom{j}{l}^{1/2} \phi_{j-l} + \mu_t
\]
\begin{align}
\gamma_t & = \sum_{j=l}^{K} \beta_j 2^{-j} \left( \frac{j}{l} \right)^{1/2} \hat{\phi}_{j-l} \\
\hat{\phi}_j & = \sum_{i=0}^{j} a_{ij} E \left[ u_i^t \right] 2^j \\
\mu_t & = \epsilon_t + \sum_{j=0}^{K} \beta_j \sum_{l=0}^{j} 2^{-j} \left( \frac{j}{l} \right)^{1/2} p_{l} \left( w_t \sqrt{2} \right) \left( p_{j-l} \left( u_t \sqrt{2} \right) - \hat{\phi}_{j-l} \right)
\end{align}

(a.4) is the recurrence formula given in the text. (a.5) defines the "nuisance parameters" as linear combinations of the uncentered moments of the error variable \( u_t \). Lastly, under (3.24) and (3.25), we have that \( E[\mu_t | z_t] = 0 \). The assumption of independence of \( z_t \) and \( u_t \) is crucial for this result. One can check that

\begin{align}
\gamma_K = \frac{\beta_K}{2^{\frac{K}{2}}}; \quad \gamma_{K-1} = \frac{\beta_{K-1}}{2^{\frac{K-1}{2}}}
\end{align}

This implies that instrumental variable is consistent in presence of measurement error, when the true specification is quadratic or linear.

**a.2 Derivation of (3.35):**

Turning to the auxiliary regression,

\begin{align}
\epsilon_t k_t & = \sum_{j=0}^{K} \beta_j p_j (w_t + u_t) (w_t + u_t) + \eta_t \epsilon_t + (w_t + u_t) \epsilon_t \\
& = \sum_{j=0}^{K} \beta_j \left( \sqrt{j + 1} p_{j+1} (w_t + u_t) + \sqrt{j} p_{j-1} (w_t + u_t) \right) + \eta_t \epsilon_t + (w_t + u_t) \epsilon_t \\
& = \sum_{j=0}^{K-1} p_j (w_t + u_t) \left( \sqrt{j} \beta_{j-1} + \sqrt{j + 1} \beta_{j+1} \right) \\
& \quad + \beta_K \sqrt{K + 1} p_{K+1} (w_t + u_t) + \beta_{K-1} \sqrt{K} p_K (w_t + u_t) + \eta_t \epsilon_t + (w_t + u_t) \epsilon_t
\end{align}

157
= \sum_{j=0}^{K-1} \left( \sqrt{j} \beta_{j-1} + \sqrt{j+1} \beta_{j+1} \right) \sum_{l=0}^{j} 2^{-\frac{4}{l}} \left( \frac{j}{l} \right)^{1/2} p_l \left( w_t \sqrt{2} \right) p_{j-l} \left( v_t \sqrt{2} \right) \\
+ \beta_K \sqrt{K + 12^{-K+1}} \sum_{l=0}^{K+1} \left( \frac{K+1}{l} \right)^{1/2} p_l \left( w_t \sqrt{2} \right) p_{K+1-l} \left( v_t \sqrt{2} \right) \\
+ \beta_{K-1} \sqrt{K2^{-K+1}} \sum_{l=0}^{K} \left( \frac{K}{l} \right)^{1/2} p_l \left( w_t \sqrt{2} \right) p_{K-l} \left( v_t \sqrt{2} \right) + \eta_t \epsilon_t + (w_t + v_t) \epsilon_t \\
= \sum_{l=0}^{K-1} p_l \left( w_t \sqrt{2} \right) \sum_{j=1}^{K-1} \left( \sqrt{j} \beta_{j-1} + \sqrt{j+1} \beta_{j+1} \right) 2^{-\frac{4}{l}} \left( \frac{j}{l} \right)^{1/2} \phi_{j-l} \\
+ \sum_{l=0}^{K+1} p_l \left( w_t \sqrt{2} \right) 2^{-K+1} \left( \frac{K+1}{l} \right)^{1/2} \beta_K \sqrt{K + 1} \phi_{K+1-l} \\
+ \sum_{l=0}^{K} p_l \left( w_t \sqrt{2} \right) 2^{-K} \left( \frac{K}{l} \right)^{1/2} \beta_{K-1} \sqrt{K} \phi_{K-l} + \sigma_{\epsilon_t} + \sigma_{\epsilon_v} + \omega_t \\
= \sum_{l=0}^{K-1} p_l \left( w_t \sqrt{2} \right) \\
\left( \sum_{j=1}^{K-1} \left( \sqrt{j} \beta_{j-1} + \sqrt{j+1} \beta_{j+1} \right) 2^{-\frac{4}{l}} \left( \frac{j}{l} \right)^{1/2} \phi_{j-l} \\
+ 2^{-K} \left( \frac{K}{l} \right)^{1/2} \beta_{K-1} \sqrt{K} \phi_{K-l} + 2^{-K+1} \left( \frac{K+1}{l} \right)^{1/2} \beta_K \sqrt{K + 1} \phi_{K+1-l} \right) \\
+ p_K \left( w_t \sqrt{2} \right) \beta_{K-1} \sqrt{K2^{-K+1}} + p_{K+1} \left( w_t \sqrt{2} \right) \beta_K \sqrt{K + 12^{-K+1}} + \sigma_{\epsilon_t} + \sigma_{\epsilon_v} + \omega_t \\
= \sum_{l=0}^{K+1} p_l \left( w_t \sqrt{2} \right) \delta_l + \omega_t \quad (a.8)

where the next to last equality uses \( \phi_0 = 1 \) and \( \phi_1 = E[\epsilon_t] = 0 \). We have:

\[
\delta_l = \sum_{j=0}^{K-1} \left( \sqrt{j} \beta_{j-1} + \sqrt{j+1} \beta_{j+1} \right) 2^{-\frac{4}{l}} \left( \frac{j}{l} \right)^{1/2} \phi_{j-l} \quad (a.9)
\]

\[
+ \beta_{K-1} \sqrt{K2^{-K+1}} \left( \frac{K}{l} \right)^{1/2} \phi_{K-l} + \beta_K \sqrt{K + 12^{-K+1}} \left( \frac{K+1}{l} \right)^{1/2} \phi_{K+1-l}
\]

for \( 1 \leq l \leq K - 1 \),

158
\[
\delta_K = \beta_{K-1}\sqrt{K}2^{-\frac{K}{4}} \\
\delta_{K+1} = \beta_K\sqrt{K+12}2^{-\frac{K+1}{4}}
\]

and

\[
\omega_t = \sum_{j=0}^{K-1} \left( \sqrt{j}\beta_{j-1} + \sqrt{j+1}\beta_{j+1} \right) \sum_{l=0}^{j} 2^{-\frac{l}{4}} \left( \frac{i}{l} \right)^{1/2} p_l \left( w_t\sqrt{2} \right) \left( p_{j-l} \left( v_t\sqrt{2} \right) - \phi_{j-l} \right) \\
+ \beta_K \sqrt{K+12}2^{-\frac{K+1}{4}} \sum_{l=0}^{K+1} \left( \frac{K+1}{l} \right)^{1/2} p_l \left( w_t\sqrt{2} \right) \left( p_{K+1-l} \left( v_t\sqrt{2} \right) - \phi_{K+1-l} \right) \\
+ \beta_{K-1}\sqrt{K}2^{-\frac{K}{4}} \sum_{l=0}^{K} \left( \frac{K}{l} \right)^{1/2} p_l \left( w_t\sqrt{2} \right) \left( p_{K-l} \left( v_t\sqrt{2} \right) - \phi_{K-l} \right) \\
+ (\eta_t e_t - \sigma_{\eta}) + ((w_t + v_t) e_t - \sigma_{ev})
\]

and \(E[\omega_t|z_t] = 0\), given (3.25) and (3.26).
Figure 3.1: Ss Curve

Figure 1: Ss Curve

$k_u$

$(k)_{\omega}$

$\omega$

$\bar{\omega}$

$k$

$k_i$
Figure 3.2: DM-FF Log Exchange Rate and Fundamental

Figure 2: DM-FF log Exchange rate and Constructed Fundamental Regime 12: 1/12/87 – 1/5/90
Figure 3: Interest Rate Differential, France–Germany
Regime 12: 1/12/87 – 1/5/90
Figure 3.4.1: COMPOUNDED EXCESS RETURN DIFFERENTIAL, EMS
Figure 3.4.2: COMPOUNDED EXCESS RETURN DIFFERENTIAL, NON-EMS

Figure 4.G: Compounded Excess Return Differential: GERMANY

Figure 4.F: Compounded Excess Return Differential: UK

Figure 4.H: Compounded Excess Return Differential: JAPAN
Figure 3.5: EXCHANGE RATE AND FITTED: DENMARK
Figure 3.6: EXCHANGE RATE AND FITTED: BELGIUM
Figure 3.7: EXCHANGE RATE AND FITTED: FRANCE
Figure 3.8: EXCHANGE RATE AND FITTED: NETHERLANDS
Figure 3.9: EXCHANGE RATE AND FITTED: ITALY
Figure 3.10: EXCHANGE RATE AND FITTED: NON EMS
Chapter 4

Federal Transfers,
Decentralization and the Labor Market

1 Introduction

Most economists would agree that Europe does not constitute an Optimum Currency Area as defined by Mundell (1961). With sticky prices and wages, a low mobility of labor, the nominal exchange rate allows member countries to correct transitory imbalances and to absorb idiosyncratic shocks. A fixed exchange rate, in these conditions, would impose unnecessary high costs of recession through reduction of output and unemployment.

Considering this problem, economists have paid increasing attention in the past few years to the role of federal fiscal policy in single currency areas, as a way to compensate for the lost exchange rate. Looking at the evidence from the United States, Economy (1990) advocates a fiscal union or at least a stronger federal budget for the European Community, following an argument made originally by Kenen (1969). Acting mechani-
cally in a counter-cyclical way through a reduction in taxes and an increase in transfers, a federal fiscal policy could absorb part of the asymmetric shocks affecting regions, thus lessening effectively the downward adjustment in output and the rise in unemployment. This cross-country automatic stabilizer would insure member countries against idiosyncratic shocks. However, this insurance aspect of the federal fiscal policy may generate the wrong incentives for pressure groups trying to capture federal transfers. The argument, present informally since early discussions on European fiscal policy is developed formally in Gourinchas (1993).\(^1\) In particular, it emphasizes the incentive problems associated with the labor market structure and the wage bargaining process.

The intuition underlying the analysis is straightforward. First, the wage bargaining process affects macroeconomic performance. By accepting more flexible wages, workers may reduce the likelihood of being unemployed should a negative shock occur and labor demand fall. Second, federal fiscal authorities may be unable to observe and monitor local wage negotiations, thus being unable to offer transfers contingent on the willingness of the parties to accommodate shocks. These two elements are sufficient to create a moral hazard problem: workers will try to capture federal transfers by hardening their bargaining position.\(^2\)

The spatial structure of the labor market determines the strength of this incentive problem. With fully decentralized negotiations, as in the US, competition between similar firms within the same region or country is sufficient to eliminate the incentive problem. At the opposite end of the spectrum, that is, in the case of full centralization of labor negotiations at the federal level, workers do internalize that they cannot free-ride on one another. Again, they face the correct incentives. The worst possible structure lies somewhere in between: national/regional negotiations and national/regional transfers.\(^3\)

Workers in a given country/region understand that they get the identical wages and

---


\(^2\)This problem is very general and can also arise between any region and the central government.

\(^3\)See Calmfors and Drifill (1988).
receive identical transfers. This congruence of interests makes rent seeking activities more likely. This situation would characterize a post monetary unification Europe with a strong federal budget.\textsuperscript{4}

The present chapter builds upon this initial work by considering the issue of fiscal federalism. It has been argued that decentralization of fiscal policy solves all incentive problems, since it effectively insulates member countries. Local fiscal authorities can provide some insurance by running temporary deficit and surpluses. Using in this way their intertemporal stabilizer,\textsuperscript{5} the argument goes, there would be no need for a federal fiscal authority and this would eliminate the moral hazard problem. In light of the recommendations of the Delors Rapport and the Maastricht convergence criterion, which imposes a 3% ceiling on national deficits, it is interesting to know whether decentralization would achieve any efficiency gain, compared to centralization of fiscal policy.

This runs counter to a well established proposition in the early literature on fiscal federalism whereby the stabilization role of fiscal policy should be centralized.\textsuperscript{6} Decentralization creates fiscal externalities between competing local fiscal authorities, leading to inefficient decisions. If the bond market is imperfect—for instance if all countries in the federation face the same interest rate—, there is an incentive for local governments to run too large deficits and to capture an excessive share of the federation's savings.\textsuperscript{7} Thus, decentralization might replace the moral hazard problem by a tragedy of the commons.

We show in this chapter that this is not quite the case. If agents do have access to international capital markets at the same conditions as local governments, there is no gain in decentralizing fiscal policy: workers can smooth optimally their consumption stream. Decentralization is totally irrelevant and ineffective to our problem. In particular, it is

\begin{itemize}
\item \textsuperscript{4}One can argue that it characterizes also post-reunification Germany as well as Russia in its relation to provinces and regions.
\item \textsuperscript{5}As opposed to intratemporal insurance through the federal transfer scheme.
\item \textsuperscript{6}See Musgrave (1971).
\item \textsuperscript{7}A simple model along these lines is presented in Canzoneri and Diba (1992).
\end{itemize}
not a substitute for federal fiscal policy.\textsuperscript{8}

If financial markets in the federation are imperfect, local fiscal policy may achieve welfare gains. This is not necessary, though. In equilibrium, access to capital markets allows local governments to insure partly their constituencies against future income fluctuations. However, borrowing introduces an asymmetry between countries, which drives them away from the First Best. In addition, borrowing amounts to a renegotiation of the federal transfer scheme. This, in turns, affects the amount of insurance provided \textit{ex-ante} by the federal fiscal authority.\textsuperscript{9} The federation might be better off without local governments.

Therefore, we do not find any substantive argument in favor of decentralization of the stabilization role of fiscal policy in a Monetary Union. We develop these ideas in details in the rest of the chapter, with a two periods version of Gourinchas (1993). Section 2 describes the model. Section 3 solves for the optimal federal fiscal policy under centralized and decentralized negotiations. With centralized negotiations, federal fiscal policy provides full insurance and achieves the first best. It is then shown that the optimal federal policy under decentralized negotiations provides some but not full insurance. Section 4 introduces local government. We show that local fiscal policy may increase welfare. However this is neither necessary nor a substitute for federal fiscal policy. Lastly, we demonstrate, as a straightforward application of Ricardian equivalence, that local fiscal policy is only effective as long as agents have imperfect access to financial markets.

\textsuperscript{8}We focus only on the stabilization role of fiscal policy. Of course, the provision of local public goods requires some decentralization of fiscal policy.

\textsuperscript{9}Although renegotiation always leads to \textit{ex-post} efficiency here, as usual.
2 The model:

The economy consists of two identical countries and lasts two periods. All foreign variables will be starred. In each country, we have a continuum of identical workers on $[0, 1]$. Each worker supplies inelastically one unit of labor in each period. There is a unique consumption good, produced in both countries. Labor is the unique factor of production. The production function $f(l, \epsilon)$ (resp. $f(l, \epsilon^*)$ in the foreign country) is concave in labor $l$. $\epsilon$ and $\epsilon^*$ represent i.i.d. idiosyncratic shocks to the domestic and foreign economy respectively, distributed according to $\Phi(\epsilon; e)$ on $[\bar{\epsilon}, \hat{\epsilon}]$, where $e$ represents the "effort" level.\textsuperscript{10} We assume that $f_{\epsilon}$ and $f_{l, \epsilon}$ are strictly positive, where $f_i$ represents the partial derivative of $f$ with respect to argument $i$. This implies that a higher value of $\epsilon$ increases both production and the marginal productivity of labor. "Effort" here measures the willingness of workers to accommodate productivity shocks. Through increased effort, workers stochastically raise their productivity, and may be able to offset partially negative productivity shocks, thus commanding a higher wage. Formally, we assume that $\Phi(\epsilon; e)$ satisfies First Order Stochastic Dominance: $\Phi_{\epsilon}(\epsilon; e) < 0$. Wages are only one, albeit important, element of the labor negotiation process. Workers and firms often bargain about working conditions, the organization of labor, the amount of monitoring etc... All these aspects presumably affect labor productivity and may be difficult to observe from the outside. Thus, our effort measure should be thought of as a catch-up variable for all the relevant characteristics, besides the wage, affecting potentially the performance of the firm. From the outsider's point of view, the firm's performance reflects a combination of the productivity shock and the effort level.

There is no migration of labor, thus the labor market clears in the usual way in each country, determining the equilibrium wage once the productivity shocks have been revealed:

\textsuperscript{10}Assuming serial independence between $\epsilon$ and $\epsilon^*$ allows us to truly focus on transitory shocks.
\[ w = f_1(1, \epsilon) \quad ; \quad w^* = f_1(1, \epsilon^*) \]

Since wages adjust after the shock is revealed, there is no unemployment in this economy. We assume that workers must set their effort variable \textit{ex-ante}. A change in effort will only change productivity from now on, and does not affect past productivity. Once the shock is realized, it is too late: workers are paid their marginal product and no degree of accommodation is possible. It is important to note that the wage in each country does not depend directly on the foreign shock or effort.

From the labor market equilibrium condition we derive the cumulative distribution of the wage \( w \) as a function of effort: \( \Phi(w; e) \), where \( w \) lies in the bounded support \([w, \bar{w}]\). Denoting \( \varphi(w; e) \) a Radon Nykodym derivative of \( \Phi(w; e) \), we assume that \( \varphi(w; e) \) satisfies the Monotone Likelihood Ratio Property.

\textbf{Assumption 2.1 (MLRP)}

\[ \frac{\partial}{\partial w} \left( \frac{\varphi}{\Phi} (w; e) \right) \geq 0 \]

When \textit{MLRP} is satisfied, higher values of realized wages are indicative of higher effort. Thus with transfers contingent on wages, effort will be monotonous in the compensation level.

In addition, we assume that \( \Phi(w; e) \) satisfies the Convexity of the Distribution Function Condition (CDFC) assumption:

---

\(^{11}\)Identical results are obtained if we introduce wage rigidities. Then, adjustment occurs through unemployment.

\(^{12}\)\( \bar{w} \) is defined by \( \bar{w} = f_1(1, \bar{\epsilon}) \), and \( w \) similarly.

\(^{13}\)Given that \( f_{1, \epsilon} > 0 \), it is equivalent to assume that \( \varphi(w; e) \) satisfies MLRP.

\(^{14}\)See Holmström (1979). Note also that \textit{MLRP} is stronger than First Order Stochastic Dominance.
Assumption 2.2 (CDFC)

\[ \Phi_{ee}(w; e) \geq 0 \]

The cumulative distribution function is convex in effort. This condition ensures that the First Order Approach is valid, so that we can replace the worker's maximization program with its first order condition.\textsuperscript{15}

Workers maximize the expected present discounted value of utility:

\[ U = E [u(c_1) - \psi(e_1) + \beta(u(c_2) - \psi(e_2))] \]

(2.1)

where \( u \) is a strictly concave Von Neumann Morgenstern utility function in consumption \( c \) and \( \psi(e) \) represents the disutility of effort \( e \). \( \beta \) is the discount factor. We assume that \( \psi \) is strictly convex in effort.

The federal government tries to maximize the expected sum of workers’ utilities in both countries, \( U + U^* \). It provides some insurance by taxing income in both countries and redistributing the proceeds.\textsuperscript{16} We suppose that these transfers are made directly to the agents in each economy, not to national governments, nor to firms. More specifically, the federal government chooses a linear tax rate \( \tau \) and transfers \( z \), at the beginning of each period, and must satisfy its intertemporal budget constraint:

\[ \tau_1 (w_1 + w_1^*) + \tau_2 (w_2 + w_2^*) = 2 (z_1 + R z_2) \]

(2.2)

where \( R \) is the equilibrium gross interest rate. We implicitly assumed that the tax rate \( \tau \) is not state-dependent. This rules out \textit{ex-ante} redistributive motives. Relaxing this assumption doesn’t change the results since we have perfect symmetry: the federal

\textsuperscript{15}See Rogerson (1985). With CDFC and MLRP, the program of the federal government is globally concave.

\textsuperscript{16}We will assume from now on that taxation is non-distortionary. With distortiomy taxation, the federal government would not offer full insurance and this would obscure more important results.
government would always choose a state-independent tax rate at the optimum.\textsuperscript{17} Local
governments, maximizing welfare in their country, can decide to run a deficit or surplus \(d\), once the state of the economy is revealed, taxing income in the second period to
reimburse, at rate \(\tau_d\).\textsuperscript{18} The budget constraint for local fiscal authorities is thus:

\[ R d = \tau_d w_2 \quad (2.3) \]

Thus, the budget constraints for domestic workers are:

\[
\begin{align*}
  c_1 + b &= w_1 (1 - \tau_1) + z_1 + d \\
  c_2 &= w_2 (1 - \tau_2 - \tau_d) + z_2 + R b 
\end{align*}
\quad (2.4) \quad (2.5)
\]

where \(b\) represents bond demand by workers in the first period. A similar set of
budget constraints holds for foreign workers.

Federal fiscal policy insures workers against bad productivity realizations, by tying
their net wages to the foreign wages. This insurance component affects domestic and
foreign workers’ incentives to exert effort \textit{ex-ante}. Intuitively, transfers reduce the income
differential between the good and the bad states. Workers are less willing to accept a
high effort since they can free-ride on their neighbors. In equilibrium however, this is
likely to give an inefficiently low effort level in both countries. Workers fail to internalize
the impact of their effort decision on foreign workers net wages.

The next section solves for the optimal efforts and tax rates when fiscal policy is fully
centralized and workers do not have access to capital markets. We then turn to the issue
of decentralization.

\textsuperscript{17}With state dependent tax rates we would have \textit{ex-ante} redistribution between the two countries.
This is optimal if the countries are not identical.
\textsuperscript{18}In this model in which all agents in a given country are identical, the only aim of local governments
is to smooth income fluctuations.
3 Centralized Fiscal Policy

We assume in this section that workers have no access to capital markets and that fiscal policy is fully centralized. This is the strongest and simplest form of imperfection of the financial markets. No intertemporal smoothing is possible. Under these assumptions, the two periods are independent, the federal government runs a period by period balanced budget and workers must consume every period their net wage: \( c_t = w_t (1 - \tau_t) + z_t \). In the remaining of this section, we drop the time subscripts. These assumptions make a stronger case for a federal fiscal policy since workers have no instrument to smooth out wage fluctuations.

3.1 Centralized Negotiations and First Best

As is usual, we first characterize the optimal policy under cooperation. This corresponds to a shift in bargaining from the national to the federal level. Since both countries are \textit{ex-ante} identical in each period, effort levels should be equal. Workers fully internalize the externality and choose \textit{ex-ante} a common effort level. The optimal fiscal policy, freed from the hidden action problem, entails full insurance for each worker: \( \tau = 1 \) and \( z = \frac{1}{2} (w + w^*) \). Consequently, workers in both countries choose \( e \) so as to satisfy:

\[
e \in \arg\max \int \int u \left( \frac{w + w^*}{2} \right) \varphi(w^*; e) \varphi(w; e) \, dw \, dw^* - \psi(e)
\]

Where we substituted the federal budget constraint (2.2) in the worker's one. We assume that this program is strictly concave in effort so that the necessary conditions are also sufficient.\(^{19}\) The first order condition for this maximization program is:

\[
2 E \left[ u \left( \frac{w + w^*}{2} \right) \frac{\varphi_e}{\varphi} \right] = \psi' \left( e \right)
\]

\( ^{19} \text{The assumption can be stated formally as: } \frac{\partial^2 E[u(e)]}{\partial e^2} - \psi(e) < 0. \)
The left hand side represents the marginal benefit of a higher effort level. Under assumption 2.1, this term is positive. Increasing $e$ by one unit shifts the distribution of productivity shocks to the right, which increases the likelihood of high wages. This increases expected utility. The right hand side measures the marginal cost of a higher effort level. We define $e_f$ as the solution for this program. It is immediate to see that this solution corresponds to the First Best, i.e. the choice of effort and transfers chosen by the Central Planner. Thus we have the first proposition.

Proposition 4.1 Centralized wage bargaining and federal fiscal policy achieve the First Best.

Under these assumptions, a federal fiscal policy is clearly optimum: it smoothes out nominal income fluctuations over space, opens the only missing market and maximizes ex-ante wages in the federation. It is important to note that our result does not depend on financial market imperfection. Since workers in both countries get the same net wages, there is no need for borrowing. Thus, even if access to capital market was granted, or even in presence of local fiscal authorities, the optimal policy would entail full insurance at the federal level. This result suggests that decentralized and centralized fiscal policy may not be perfect substitutes. Although we are limiting ourselves to the case of centralized wage bargaining, federal fiscal policy achieves the first best by offering full insurance, period per period. By comparison, local fiscal policy cannot offer full insurance, since intertemporal insurance schemes create ex-post heterogeneity between countries.²¹

3.2 Decentralized Bargaining

In this subsection, we assume that the wage bargaining occurs at the national, as opposed to federal, level. Thus, workers in the domestic country take the foreign degree of effort

---
²⁰ See the appendix for a proof.
²¹ Optimal insurance requires that the ratio of marginal utilities of different agents stay constant over time. This condition is not satisfied when agents insure themselves through intertemporal borrowing.
as given. Workers in both countries play a Nash game in both periods. In addition, we assume that $e$ and $e^*$ are not observable at the federal level while the state of nature (i.e. wages) is both observable and verifiable. This last assumption has the following interpretation. While wages and employment decisions are fairly easily observable, labor negotiations often encompass other aspects of the employee-employer relationship. Benefit policies, the degree of autonomy left to employees, the organization of the firm (layers of supervision, monitoring activities...), career opportunities inside the firm as well as promotion policies, are very relevant characteristics of the working environment. Thus the outcome of a given round of negotiations may be poorly summarized by the resulting wage. In addition, different countries or regions have different institutional characteristics that make direct comparison based solely on wages difficult. If only for this last reason, local authorities may have better information than federal authorities. Thus, the federal authority faces a standard moral hazard problem. Workers will try to capture federal transfers by hardening their bargaining position and exerting less effort. We assume that the timing within each period is as follows:

- at the beginning of the period the federal government announces the federal tax rate set for the period;

- then, workers in each country play a Nash game, choosing their effort level taking as given the foreign one. Workers in both countries also internalize the federal and local budget constraint;

- lastly the state of both economies is revealed. The wage clears the labor market. Federal transfers are residual.

---

22 The cost of gathering information on the willingness of workers to accommodate shocks may simply be extremely high. If the wage negotiation occurs at the national level, local governments may have a fair idea of the worker's effort. However, if they share their constituencies' interests, they have no incentives to report the truth to the federal government.
We focus on the symmetric Nash subgame perfect equilibrium. Given that the effort level is chosen \textit{ex-ante}, we can rewrite expected utility as:

\[ U = E[u(c)] - \psi(e) \]

The per period federal budget constraint is:

\[ \tau(w + w^*) = 2z(w, w^*) \quad \forall (w, w^*) \tag{3.7} \]

where \( z \) represents the transfers received by workers in each country. We solve first the Nash equilibrium between the two countries. Each country solves the following maximization program:

\[ e \in \arg \max E[u(c)] - \psi(e) = \int \int u(c) \varphi(w^*; e) \varphi(w; e) \, dw \, dw^* - \psi(e) \]

\[ \text{s.t.} \quad c(w, w^*) \leq w - \frac{\tau}{2} (w - w^*) \]

where we substituted the federal budget constraint (3.7) in the workers' budget constraint. Workers' expected utility depends on foreign workers' wages through \( \varphi(w^*; e) \). Net transfers are positive when the domestic wage is lower than the foreign wage. We assume that this program is strictly concave in effort so that the necessary conditions are also sufficient.\textsuperscript{23} The first order condition (F.O.C.) of this program is:

\[ \int \int u \left( w - \frac{\tau}{2} (w - w^*) \right) \varphi(w^*; e) \varphi_e(w; e) \, dw \, dw^* = \psi'(e) \tag{3.8} \]

We can rewrite this condition in the following form, easier to interpret:

\textsuperscript{23}The assumption can be stated formally as: \[ \frac{\partial^2 E[u(c)] - \psi(e)}{\partial e^2} < 0 \] or \[ \int \int u(c) \varphi(w^*; e) \varphi_{ee}(w; e) \, dw \, dw^* - \psi''(e) < 0. \]


\[ E \left[ u(c) \frac{\varphi_e}{\varphi} \right] = \psi'(e) \] (3.9)

Given that the score vector \( \frac{\varphi_e}{\varphi} \) is centered (i.e. \( E \left[ \frac{\varphi_e}{\varphi} \right] = 0 \)), the left hand side is the covariance between \( u(c) \) and the score vector. Under assumption 2.1, this covariance term is positive and represents the marginal benefit of an increase in effort: increasing effort shifts the distribution of shocks to the right. This increases expected utility.\(^{24}\) The right hand side represents the marginal cost. The solution to this F.O.C. defines implicitly a Best Response for domestic workers:

\[ e = B(e^{*}, \tau). \]

We show in the appendix that \( \frac{\partial B}{\partial e} \leq 0 \). Domestic and foreign effort are strategic substitutes. An increase in the foreign effort level decreases domestic effort. Looking at (3.8), we see that the interaction between \( e \) and \( e^{*} \) comes through federal transfers. When \( \tau = 0 \), domestic and foreign country are decoupled and set their effort level independently.

In addition, we assume that effort decreases with the tax rate.\(^{25}\)

**Assumption 3.1**

\[ \frac{\partial B(e^{*}, \tau)}{\partial \tau} < 0 \]

This case is the more interesting for our analysis, since we expect a higher federal tax rate to reduce the cost of choosing a low effort and ending up with a low wage. With this assumption, in hand, we can solve for the symmetric Nash equilibrium played by workers in both countries:

\(^{24}\)For a proof, see the appendix.

\(^{25}\)This is not guaranteed, see the appendix for a discussion.
\[ e = e^* = B(e, \tau) \equiv B(\tau) \quad (3.10) \]

with \( B'(\tau) < 0 \). This is the Incentive Compatibility (IC) constraint faced by the federal government. Workers in both countries fail to internalize the impact of their effort decision on foreign net wages. By decreasing their effort level, they increase the likelihood of receiving positive net transfers. This decreases expected foreign net wages. We follow the First Order Approach and solve for the optimal tax rate by maximizing the federal government objective function under the IC constraint (3.10).\(^{26}\) The First Order Approach is valid, under MLRP and CDFC. Formally, the federal government solves:

\[
\begin{align*}
\tau \in \arg\max & \quad E[u(c) + u(c^*)] - \psi(e) - \psi(e^*) \\
\text{s.t.} & \quad c(w, w^*) \leq w - \frac{\tau}{2}(w - w^*) \\
& \quad c^*(w, w^*) \leq w^* + \frac{\tau}{2}(w - w^*) \\
& \quad e = e^* = B(\tau)
\end{align*}
\]

Writing the F.O.C. and substituting (3.10), using the symmetry of the model, the optimal federal tax rate \( \tau \) satisfies:\(^{27}\)

\[
E \left[ u(c) \frac{\varphi_e}{\varphi_e^*} \right] B'(\tau) - E \left[ u'(c) \frac{\Delta w}{2} \right] = 0 \quad (3.11)
\]

where \( \Delta w = w - w^* \).

\(^{26}\)We should also add an Individual Participation constraint: in equilibrium, workers from both countries must be willing to participate in the federal fiscal scheme. For simplicity, we assume that workers cannot decide not to participate. Reservation utility is 0 and will not bind in equilibrium.

\(^{27}\)See the appendix. In particular, we show that the first term is negative when assumption 2.1 is satisfied, while the second is positive.
Let start with the interpretation. Increasing \( \tau \) has two effects. First, by decreasing \( e \) and \( e^* \), it increases the probability of recession in both countries (first term) and decreases expected utility. This is the incentive effect. Second, it transfers income between countries (second term). This is the insurance effect. We define \( \tau_n \) and \( e_n = B(\tau_n) \) respectively as the optimal tax rate and the associated optimal effort level – where the subscript \( n \) stands for Nash equilibrium. Comparing (3.8) with (3.6), we see that the social marginal benefit is twice the private one, while the social marginal cost is unchanged. Hence it is immediate that \( e_f > e_n \). Two important results are summarized in the following proposition.

**Proposition 4.2** In the symmetric subgame perfect Nash equilibrium (SPNE), without local governments and without access to capital markets:

(i) workers in each country are not fully insured: \( \tau_n < 1 \)

(ii) workers in each country are partly insured: \( \tau_n > 0 \)

(iii) the effort level is lower than without transfers: \( e_n \leq B(0) \)

**Proof.** the results are immediate.

(i) is shown by contradiction. Suppose we have full insurance. Then, \( c = \frac{w + w^*}{2} \). We can rewrite the second term of (3.11) as:

\[
E \left[ u'(c) \frac{\Delta w}{2} \right] = \int \int u'\left(\frac{w + w^*}{2}\right) \frac{w - w^*}{2} \varphi(w; e) \varphi(w^*; e) \, dw \, dw^*
\]

\[
= \int \int u'\left(\frac{w + w^*}{2}\right) \frac{w - w^*}{2} \varphi(w^*; e) \varphi(w; e) \, dw \, dw^*
\]

\[
= -E \left[ u'(c) \frac{\Delta w}{2} \right] = 0
\]

where the second lines comes from permuting \( w \) and \( w^* \). Thus the marginal benefit is 0. The first term is:
\[
E \left[ u' \left( \frac{w + w^*}{2} \right) \varphi_e \left( w^* ; e \right) \varphi \left( w ; e \right) \right] dw dw^* \\
= \int \int u \left( \frac{w + w^*}{2} \right) \varphi_e \left( w ; e \right) \varphi \left( w^* ; e \right) dw dw^* \\
= E \left[ u' \left( \frac{\varphi_e}{\varphi} \right) \right] = \psi' \left( e \right) > 0
\]

where the second line comes from permuting \( w \) and \( w^* \), and the last one from the worker’s F.O.C.. Thus, the necessary and sufficient condition (3.11) is not satisfied when we have full insurance. From the concavity of the program, we know that the solution entails a lower tax rate.

(ii) with zero transfers, the two countries are independent. In particular, the first term vanishes, as the score vector is centered. But, by lemma C.1 from the appendix, and lemma C.3, we know that the marginal benefit is positive. Thus marginal benefit exceeds marginal cost and it is optimal to provide some insurance.

(iii) as \( B' < 0, B(0) > B(\tau_n) = e_n \) where \( \tau_n > 0 \) is the second best tax rate.

Even though \( \tau_n \) is a Second Best policy (which means that it does better than the no-transfer policy in terms of expected utility), it does worse in terms of expected wages than no transfer at all. From a macroeconomic point of view, these results shed some pessimistic light on the role of federal transfers as a shock absorbing device. Full insurance is suboptimal, and there is a clear trade-off between insurance and incentives. This, in itself, is hardly a new result, and the argument was often stated in informal terms.\(^{28}\) There are two immediate ways to circumvent the problem. First, the federal government could monitor wage negotiations. Apart from the already mentioned high cost of doing so, this would allow the Federal institutions to implement the First Best solution. The federal government could impose a common wage level in both countries thus internalizing

the externality. The other solution would be a shift in the centralization level of wage bargaining, from the national to the federal level. Workers, facing common prospects and setting a common wage level would, on their own, internalize the incentive problem.

In practice, all wage negotiations need not be conducted at the federal level. Some degree of geographic and/or sectoral centralization of the wage bargaining is sufficient to weaken the incentive problem. In Germany, for instance, most of the wage bargaining occurs at the industry-regional level. However, a nation-wide confederation sets the tune between the industry-regional specific unions. Some unions (in the metallurgical sector for instance) also play a leading role in reaching pattern-making agreements. In the case of Europe, a representative body of European Trade Unions, located in Brussels, lobbies actively the European Commission and the Parliament on behalf of European workers. Industry wide wage negotiations should not prove too difficult to conduct in some particular sectors. The steel industry could be a case in point. Throughout Europe, steel plants have been heavily restructured, if not closed, and the spectrum of overproduction is always present. Steel workers in countries as diverse as Italy, Germany and France face the same difficulties and could join their effort at the Community level. European farmers, heavily dependents on the Common Agricultural Policy already voice their concerns in Brussels, through international demonstration, and play an active lobbying role. All this should lead to a more rapid internalization of the federal transfer scheme. Once leading sector's union will have concluded global agreements, one would expect them to spread over industries and regions. Eventually, the wage negotiation process will take place at the European level and the federal transfer scheme will play fully its role: insuring member countries against bad draws.
4 Decentralized Fiscal Policy

Nevertheless, the transition to centralized bargaining is likely to take time. Meanwhile, incentive problems will be present. In order to eliminate them, it has been argued that member countries should be allowed to run temporary deficits and surpluses. This would allow them to absorb transitory imbalances through their own intertemporal stabilizer. This would also solve the moral hazard problem since workers' effort decision would be unaffected by their neighbors. In other words, decentralization seems to solve most problems associated with the loss of exchange rate as a policy instrument. The question is interesting on several grounds. First, from a pragmatic point of view, it challenges the usefulness of the Maastricht Treaty convergence criterion and the rationale for some recommendations of the Delors Rapport. Both put a ceiling on national borrowing. Deficits are supposed not to exceed 3% of any country's GNP. It has already been argued that such a ceiling might have a severe impact on national economies, by imposing unnecessary pro-cyclical fiscal policies.

Second, the question here is one of fiscal federalism: which fiscal jurisdiction should be responsible for insuring member countries against idiosyncratic shocks? The early literature on fiscal federalism, assuming perfect information - e.g. Musgrave (1971) - gives a clear answer: if the only objective of fiscal policy is to stabilize output, it should be centralized.

Decentralization gives local authorities the wrong incentives. In particular, national governments might be tempted to run excess deficits or to overspend if their fiscal policy has a limited impact on bond rates in the Union. Financial integration creates a classic coordination problem. In the Mundell-Fleming framework, this externality is apparent. Within a monetary union, fiscal expansions attract foreign capital as the domestic interest rate increases. This reduces money supply in the foreign country and decreases foreign

---

29 See Canzoneri and Diba (1992) and Kehoe (1987). Delors (1989) warns that "access to a large capital market may... facilitate the financing of economic imbalances."
output. If this effect dominates the increased demand for foreign goods, we have a recession abroad. Thus, uncoordinated fiscal policies have the potential for "beggar-thy-neighbor" type effects.

In our case however, local fiscal policies will only be useful when agents have a limited access to capital markets. Whenever agents have access to an international credit market, decentralized fiscal policy will be totally useless: Ricardian Equivalence will hold. The private sector can reverse whatever decision is taken by local governments.\(^{30}\)

When agents cannot access capital markets, local governments, observing the state of the world, can decide to borrow today on their behalf and tax income tomorrow to repay the debt. The interest rate will then endogenously equilibrate net bond demand and supply from each country. We start by solving the optimal fiscal structure problem when agents cannot smooth wage fluctuations. This provides the strongest rationale for a local fiscal authority. We then consider briefly the case where agents have full access to capital markets.

### 4.1 Decentralized policy, imperfect capital markets.

This subsection assumes that agents do not have access to capital markets. We have both a local and a federal policy. Each fiscal authority tries to maximize welfare in its jurisdiction. This allows us to look directly for the optimal fiscal structure. Local governments can decide to issue debt once the state of the economy is revealed. This debt is held by the foreign government. Thus, fiscal policy is to be interpreted as a direct loan from one government to another. We also assume that the inter-governmental loan market is competitive, so that both local governments take the interest rate as given.

Workers in both periods choose their optimal effort level strategically, taking the foreign effort level and the federal budget constraint as given. We look for the Subgame

\(^{30}\)This is a consequence of the absence of any provision role for fiscal policy. In presence of local public goods to finance, local governments might be tempted to borrow excessively.
Perfect Nash Equilibrium (SPNE) of this game by solving sequentially, starting with the last period subgame.

last period;

Domestic workers face the following problem:

\[
e_2 \in \arg \max E [u(c_2)] - \psi(e_2) \\
\text{s.t. } c_2 = w_2 \left(1 - \frac{\tau_2}{2}\right) + w_2^* \frac{\tau_2}{2} - Rd
\]

where the subscript denotes the period, \(d\) represents the local public deficit in the first period, \(R\) the gross interest rate on the public debt determined endogenously, and \(\tau_2\) the federal tax rate in the second period.\(^3\) The F.O.C. for this program is similar to (3.9):

\[
E \left[ u(c_2) \frac{\varphi e}{\varphi} \right] = \psi'(e_2)
\]

The analysis is similar to the previous section. This condition defines implicitly a Best Response function \(e_2 = B_2(e^*_2, \tau_2, d, R)\). The partial derivatives have the following signs.\(^3\):

\[
\frac{\partial B_2}{\partial e^*_2} \leq 0 ; \quad \frac{\partial B_2}{\partial \tau_2} \leq 0 ; \quad \frac{\partial B_2}{\partial d} \geq 0 ; \quad d \frac{\partial B_2}{\partial R} \geq 0
\]

The first two partial derivatives have the same sign as in the previous section. A higher deficit \(d\) lowers expected utility in the second period. This increases expected marginal utility, and the marginal benefit from a higher effort level. Indebted countries

\(^3\)The federal budget is still balanced period by period.  
\(^3\)Under assumption 3.1. See the appendix.
will tend to exert higher effort. An increase in the interest rate depends on the net position of the country. If the country is a net debtor \((d > 0)\), an increase in interest rates increases the financial burden. This leads to a higher effort level. The Nash equilibrium is not symmetric in the second period. If \(d > 0\), the Best Response of the domestic workers is shifted upwards while the Best Response of foreign workers is shifted leftwards. Thus \(e_2 > e_2^*\). By symmetry between the two countries however, we can write:

\[
e_2 = B_2 (\tau_2, d, R) \quad \text{and} \quad e_2^* = B_2 (\tau_2, -d, R)
\]  

(4.13)

The federal government maximizes total expected utility, taking as given the equilibrium choices of effort as a function of the federal tax rate:

\[
\tau_2 \in \text{arg max} \quad E [u (c_2) + u (c_2^*)] - \psi (e_2) - \psi (e_2^*)
\]

\[
s.t. \quad c_2 \leq w_2 - \frac{\tau_2}{2} (w_2 - w_2^*) - Rd
\]

\[
c_2^* \leq w_2^* + \frac{\tau_2}{2} (w_2 - w_2^*) + Rd
\]

\[
e_2 = B_2 (\tau_2, d, R) \quad ; \quad e_2^* = B_2 (\tau_2, -d, R)
\]

The F.O.C. for this program becomes:

\[
E \left[ u (c_2) \frac{\phi_e}{\phi^*} \right] \frac{\partial B_2^*}{\partial \tau_2} + E \left[ u (c_2^*) \frac{\phi_e}{\phi} \right] \frac{\partial B_2}{\partial \tau_2} - E \left[ (u' (c_2) - u' (c_2^*)) \frac{\Delta w_2}{2} \right] = 0 \quad (4.14)
\]

The interpretation is similar to (3.11): the first two terms represents the incentive effect. Effort in both countries decrease as the federal tax rate increases. The last term represents the insurance effect. This defines an optimal federal tax rate, as a function of the debt burden \(d R: \tau_2 = \tau_2 (d R)\).

A higher debt burden has a complex effect on the federal tax rate. When \(d R\) increases,
the foreign country is _ex ante_ richer. This decreases the marginal cost of federal taxes abroad and increases the domestic one. These two effects go in opposite direction. In addition, we also have the indirect effect on effort levels. Increasing debt increases effort in the domestic country and decreases it abroad. The overall effect is ambiguous. By symmetry between the two countries, however—and simple inspection of the F.O.C.—it is immediate that the tax rate is symmetric around \( d = 0 \): \( \tau_2 (d R) = \tau_2 (-d R) \). Thus the tax schedule is flat at \( d = 0 \).

**First period.**

The local government will choose debt so as to solve:

\[
\begin{align*}
    d & \in \arg \max \{ u(c_1) + \beta E [ u(c_2) ] \} \\
    \text{s.t.} \quad c_1 & = w_1 \left( 1 - \frac{\tau_1}{2} \right) + \frac{\tau_1}{2} w_1^* + d; \\
    c_2 & = w_2 \left( 1 - \frac{\tau_2}{2} \right) + \frac{\tau_2}{2} w_2^* - R d
\end{align*}
\]

taking as given the interest rate \( R \) and next period federal tax rate \( \tau_2 \).\(^{33}\) \( \beta \) is the discount factor. The F.O.C. is the traditional Euler equation:

\[
u'(c_1) = \beta R E [ u'(c_2) ] \tag{4.15}\]

The left hand side represents the marginal benefit of one extra unit of deficit \( d \). It allows a higher consumption level. The right hand side represents the marginal cost.\(^{34}\)

\(^{33}\)By the Enveloppe Theorem, we do not need to consider the impact of debt on the second period effort level.

\(^{34}\)The local government does not take into account the impact of \( d \) on the federal tax rate. This is the case as the federal tax rate depends in equilibrium on the value of the debt, while (4.15) determines bond supply for a given interest rate. Nothing would change if we consider this interaction since it does not affect the equilibrium interest rate.

192
This determines the optimal debt level, as a function of the realized net wage level, \( \tilde{w}_1 = w_1 - \frac{T_1}{2} \Delta w_1 \), the interest rate and the first period federal tax rate: \( d = d(\tilde{w}_1, R, \tau_1) \), with:

\[
\frac{\partial d}{\partial \tilde{w}_1} < 0 \quad ; \quad \frac{\partial d}{\partial R} < 0 \quad ; \quad (\Delta w_1) \frac{\partial d}{\partial \tau_1} \geq 0
\]

The last expression is straightforward: a higher federal tax rate in the first period decreases ex-post income differential, thus reducing the need for borrowing. In this sense, debt and taxation in the first period are strategic substitutes. Borrowing and lending allow local authorities to compensate ex-post for the insurance that is not provided by the federal government. Since the federal fiscal authority will not provide full insurance precisely in order to preserve incentives, the availability of ex-post borrowing acts as a renegotiation device. Even though this ensures ex-post efficiency, it weakens further the incentive scheme that is used by the federal fiscal authority. We have already seen that with full insurance in the first period, there is no borrowing ex-post.

In addition, we notice that \( d \) doesn’t depend on the first period effort level, since the state of the economy is already revealed.

Equilibrium on the bond market implies:

\[
d(\tilde{w}_1, R, \tau_1) + d^*(\tilde{w}_1^*, R, \tau_1) = 0
\]

which determines the equilibrium interest rate \( R \). We can then write the equilibrium borrowing schedule as: \( d = d(w_1 - w_1^*, \tau_1) \). International borrowing and lending allows local governments to smooth out income fluctuations. However, it can only smooth income in case of asymmetric shocks. Otherwise, both countries would like to borrow, or lend, simultaneously the same amount. The interest rate would then rise up to the point where:
\[ u'(w_1 \left(1 - \frac{\tau_1}{2}\right) + \frac{\tau_1}{2}w_1^*) = \beta RE \left[u'(w_2 \left(1 - \frac{\tau_2(0)}{2}\right) + \frac{\tau_2(0)}{2}w_2^*)\right] \]  

(4.16)

i.e. \( d = d^* = 0 \). Conversely, when there is an ex-post asymmetry between countries, there is some scope for international borrowing and lending. Thus we have: \( \frac{\partial d}{\partial \Delta w} \) and \( d(0, \tau) = 0 \).

In the second stage of the first period workers choose their effort level and play a Nash game. Given that the effort level in the first period does not affect the debt decision, the necessary and sufficient condition is still (3.9), corrected for the presence of international borrowing:

\[ E \left[u(c_1) \frac{\phi_e}{\varphi}\right] = \psi'(e_1) \]

(4.17)

\[ c_1 = w_1 \left(1 - \frac{\tau_1}{2}\right) + \frac{\tau_1}{2}w_1^* + d(\Delta w_1, \tau_1) \]

We now state the following result:

**Proposition 4.3**: assume that \( \frac{\partial d}{\partial \Delta w} < \frac{u''}{u'} d \). Then,

(i) the effort level in the first period, for a given federal tax rate is smaller than without local borrowing;

(ii) the effort level in the second period, for a given federal tax rate, can be higher or lower depending on the debt level;

**Proof.** see the appendix. The technical condition required for the proposition ensures that the marginal utility has enough curvature compared with the borrowing policy. The curvature of the borrowing policy is likely to depend on the distribution of shocks. If the productivity shocks are very similar, there is not much scope for borrowing.
and the condition is likely to be satisfied. If instead the shocks can differ widely, then the condition might not be satisfied. The intuition for the result is the following. Ex-post borrowing and lending acts as a form of renegotiation. It allows workers to insure themselves once the effort decision is made and the state revealed. Thus it reduces the marginal benefit of effort in the first period. This shifts the distribution of wages to the left in both countries: expected wages in the first period are lower for any given federal tax rate. In the second period, effort increases with debt.

Lastly, we solve for the optimal federal tax rate in the first period. The program is:

$$\tau_1 \in \arg \max \ E \left[ u(c_1) + \psi(c_1^*) - \psi(e_1) - \psi(e_2^*) + \beta \{ u(c_1) + u(c_1^*) - \psi(e_2) - \psi(e_2^*) \} \right]$$

s.t.  
$c_1 \leq w_1 - \frac{\tau_1}{2} (w_1 - w_1^*) + d$
$c_1^* \leq w_1^* + \frac{\tau_1}{2} (w_1 - w_1^*) - d$
$e_1 = e_1^* = B_1(\tau_1) \ ; \ d = d(\Delta w, \tau_1)$

The solution for this program satisfies:

$$E \left[ u(c_1) \frac{\partial \psi}{\partial \tau_1} \right] \frac{\partial B_1^*}{\partial \tau_1} - E \left[ u'(c_1) \frac{\Delta w_1}{2} \right] = 0 \quad (4.18)$$

The solution satisfies proposition 4.2 – the proof is identical: in equilibrium, the federal government will provide some, less than full, insurance. This implies the following result:

**Proposition 4.4** Decentralization is not a substitute for a federal government. In the symmetric Subgame Perfect Equilibrium of the game, welfare is always higher in the presence of a federal government $(\tau_1, \tau_2) > 0$.

This is a partial answer to the general question of the optimal tax structure: it is never optimal to decentralize completely fiscal policy, i.e. to rely solely on local govern-
Intratemporal insurance is more efficient than intertemporal insurance since the latter creates *ex-post* asymmetries between countries. We can now ask whether it is optimal to decentralize at all. In order to do this, we must compare the welfare achieved in the federation under centralized fiscal policy, without local governments, and the welfare achieved when the two authorities coexist. This comparison is hard to conduct. Intuition suggests the following. In the first period, since effort is likely to be lower (due to the possibility to borrow), the federal tax rate is going to be lower than in the absence of borrowing: . The federal fiscal policy provides less insurance and let workers rely more on borrowing, as federal taxation and international borrowing are strategic substitutes. Welfare in the first period is likely to be relatively unchanged. However, borrowing introduces an asymmetry between countries in the second period. In other words, borrowing is not an efficient form of risk pooling. Effort decisions and federal taxation are likely to reduce this asymmetry—as effort increases when the country is indebted—, not to eliminate it. From the federation point of view, this asymmetry reduces total welfare.\(^{35}\) Federal fiscal policy, on the other hand, preserves this symmetry. Thus, we propose the following conjecture:

**Conjecture 4.1** Decentralization introduces asymmetries between countries. Borrowing countries have a lower expected total income than lending countries. This asymmetry drives welfare away from the First-Best in the second period. At the same time, international borrowing has ambiguous effects on ex-ante welfare in the first period since (a) it allows income smoothing and (b) it leads to a reduced federal redistribution. The optimal fiscal structure will depend on the strength of this three effects.

We emphasize that the result does not depend on externalities associated to decentralized fiscal policy, contrary to the early work on the topic. In a symmetric economy, the First Best entails full risk sharing and a symmetric income in every period. A federal

\(^{35}\)We have full risk sharing and symmetry in the First Best.
fiscal policy does not introduce any asymmetry between countries, while international borrowing introduces ex-post asymmetries. Thus, centralization is likely to be preferred. This conjecture, if valid, suggests another rationale for the Delors and Maastricht deficit requirements. If local fiscal policy can reduce welfare through its renegotiation effect, it is possible to increase welfare by reducing the scope for local fiscal policy. Deficit ceilings could prevent local government from borrowing too heavily on the European bond market. Although the argument is theoretically appealing, it is only valid if there exists a large enough federal budget to compensate for the lost fiscal autonomy. Since federal and local fiscal policies are strategic substitute, a decrease in local borrowing calls for an increase in federal redistribution. In the absence of any substantive federal policy, as is the case in Europe now, imposing deficit ceilings only worsens macroeconomic outcome by ruling out any insurance. This is hardly an optimal policy.

4.2 Decentralized fiscal policy, perfect credit market

We assumed so far that agents could not borrow or lend in the Federation capital market. This assumption made the strongest case for decentralization. Nonetheless, we showed that decentralized fiscal policy is never optimal unaccompanied by federal taxes and transfers, and that it is likely to be sub optimal. Turning now to the case of perfect financial markets, we get the following immediate result as a trivial application of Ricardian equivalence. With perfect capital markets, the path of public debt is irrelevant, as long as the intertemporal budget constraint of the fiscal authority is satisfied.

**Proposition 4.5** With perfect access to capital markets, decentralized fiscal policy is ineffective and irrelevant.

**Proof.** with perfect capital markets, domestic workers have the following budget constraints, substituting the local government budget constraints:
\[ c_1 + b - d = w_1 (1 - \tau_1) + z_1 \quad \text{in the first period} \]
\[ c_2 = w_2 (1 - \tau_2) + z_2 + R (b - d) \quad \text{in the second period} \]

Thus the choice variable is net savings \( b - d \). Equilibrium on the credit market implies the equality of bond demand and supply: \( b + b^* = d + d^* \), or \( (b - d) + (b^* - d^*) = 0 \) which doesn't depend on \( d \) and \( d^* \) directly. Hence, the borrowing decisions of the local governments are totally irrelevant. The results of the previous subsection apply. \( \square \)

In particular, even with perfect capital markets, it is optimal to preserve a federal fiscal policy. It is important to emphasize what this last result tells us about the optimal fiscal structure of a monetary union. While Ricardian equivalence holds trivially when agents have perfect access to capital markets, proposition 4.5 provides a metric with which to evaluate the need for local government. In other words, local governments may be useful only if they can get access to financial liquidities on better terms than the constituencies they represent.

5 Conclusion:

We have presented a model of moral hazard generated by the insurance component of a federal transfer scheme. These incentives effects are likely to depend heavily on the labor market structure. As a general result, familiar in the literature on insurance and incentives, it is not possible to achieve full insurance.

We then investigated whether decentralized fiscal policy could achieve welfare gains. Local governments can borrow and lend in order to smooth consumption stream in their jurisdiction. The answer depends on the access to capital markets. If agents can borrow and lend freely at the same rate as the local governments, local governments have are ineffective and irrelevant. On the contrary, when access to capital markets is imperfect, local governments may achieve welfare gains by stabilizing consumption streams and providing insurance beyond private access to capital market. This may lead to a lower welfare if it induces a large drop in federal insurance. Moreover, this creates asymmetries.
between countries and is likely to reduce future welfare. More importantly, we show that it is never optimal to decentralize totally fiscal policy. The federation is always better off when a federal government offers some insurance to member countries. Although our results suggest that local fiscal policy should be limited, it also implies that a large federal fiscal policy must be ready to absorb idiosyncratic shocks. In the absence of the latter, any restriction on the ability of countries to smooth income intertemporally effectively rules out any insurance. This is far from optimal.

It is evident that these results apply only to the insurance component of fiscal policy. In particular, we ruled out any demand effect, coming from multiplier-like mechanisms. We also assumed that local government provide only insurance. There are some well-known arguments in favor of decentralization of the provision role of fiscal policy. However, the results presented infirm some presumptions regarding the benefits of decentralization in solving the incentive problem faced by the federal government.

Federal fiscal policy is not the only way to cope with asymmetric shocks. Migration, by equalizing returns to labor in different countries, provide some cushioning. We leave to future work the task of investigating the relationship between federal transfers and migration. In particular, migration is a strong source of externalities. We suspect that a federal government trying to limit migration has no choice but to increase federal insurance, thus worsening the effort level in the federation.
Appendix C

We start with a proof that higher effort level increases expected utility of consumption.

Lemma C.1 if \( f(x) \) and \( g(x) \) are two non-decreasing continuous functions, with zero mean, then \( E[f(x)g(x)] \geq 0 \) where \( x \) is distributed according to \( \varphi(x) \) on \([a,b]\). If \( f \) and \( g \) are strictly increasing, with zero mean then \( E[f(x)g(x)] > 0 \).

Proof. \( E[f(x)g(x)] = \int \int f(x)g(x)\varphi(x)\,dx \geq g(a)\int f(x)\varphi(x)\,dx = 0 \). The second part of the proof is similar. \( \square \)

We apply lemma C.1 to the left hand side of (2.1), using \( u' > 0 \) and assumption 2.1.

Lemma C.2 \( \frac{\partial R}{\partial e^*} \leq 0 \), and \( \frac{\partial R}{\partial x} \) has an indeterminate sign.

Proof. differentiating totally the worker's F.O.C. (3.9), we get:

\[
\left\{ E\left[u(c)\frac{\varphi_e}{\varphi}\right] - \psi''(e) \right\} \, de + E\left[u(c)\frac{\varphi_e}{\varphi} \frac{\varphi_e}{\varphi^*}\right] \, dc^* - E\left[u'(c)\frac{\varphi_e}{\varphi} \frac{\Delta w}{2}\right] \, dr = 0
\]

the first term on the left hand side (in parenthesis) is negative when the program is concave: this is simply the second order condition. The second term is negative. To show this, we apply twice lemma C.1 by rewriting

\[
E\left[u(c)\frac{\varphi_e}{\varphi} \frac{\varphi_e}{\varphi^*}\right] = \int \left( \int u(c)\frac{\varphi_e}{\varphi} \, d\Phi \right) \frac{\varphi_e}{\varphi^*} \, d\Phi^*
\]

From lemma C.1, we know that the sign of this expression depends on the derivative of the inner integral with respect to \( w^* \). But
\[
\frac{\partial \int u'(c) \frac{\varphi_e}{\varphi} d\Phi}{\partial \omega^*} = \frac{\tau}{2} \int u'(c) \frac{\varphi_e}{\varphi} d\Phi
\]

and reapplying lemma C.1 to this expression, we know that it is negative, as \(u'' < 0\).

The third term has an ambiguous sign. Increasing the tax rate reduces the wage differential between countries. This affects the marginal benefit. If the marginal benefit increases, workers will be willing to provide more effort. If the marginal benefit decreases, workers will decrease their effort level. We can decompose this term as follows:

\[
-E \left[ u'(c) \frac{\varphi_e \Delta w}{\varphi} \right] = \int \left( \int \dot{u}(c) \frac{\varphi_e}{\varphi} d\Phi \right) \frac{\omega^*}{2} d\Phi - \int \left( \int \dot{u}(c) \frac{w \varphi_e}{2 \varphi} d\Phi \right) d\Phi^*
\]

The first term on the right hand side is negative, as the inner integral is negative by application of lemma C.1. When the federal tax rate increases, workers receive higher transfers when the state of the domestic economy is bad. This decreases marginal utility and tends to decrease the effort level. This is the effect we are interested in. The second term on the right hand side has an ambiguous sign. Formally, \(-\frac{\partial u(c) w}{\partial w} = -u''(c) w - u'(c)\). The sign is ambiguous, so that lemma C.1 does not apply. The first term represents the loss in utility when the tax rate increases. Net transfers are negative when the domestic wage is higher than the foreign one. The last term represents the decrease in expected wage.

In the rest of the chapter, we assume that the fall in marginal utility when the tax rate increases dominates the other effect, so that \(\frac{\partial B}{\partial \tau} < 0\).

**Lemma C.3** \(E \left[ u(c) \frac{\varphi_e}{\varphi} \right] B'(\tau) \leq 0 \ ; \ -E \left[ u'(c) \frac{\Delta w}{\omega^*} \right] \geq 0\)

**Proof.** the first part of the proof is now familiar. It is a simple application of lemma C.1, using \(B' < 0\). The second part is almost as trivial. Define \(v = \Delta w\), the wage differential. Then

\[
-E \left[ u'(c) \frac{\Delta w}{2} \right] = \int \left( \int u'(w - \frac{\tau}{2} v) \frac{\varphi(v)}{2} dv \right) \varphi(w) dw
\]

Applying lemma C.1 to the innermost integral, we have the result.

**Lemma C.4** \(\frac{\partial B_2}{\partial \omega^2} \leq 0 \ ; \ \frac{\partial B_3}{\partial \tau^2} \leq 0 \ ; \ \frac{\partial B_4}{\partial d} \geq 0 \ ; \ \frac{\partial B_5}{\partial R} \geq 0\).
**Proof.** for the first two derivatives, see lemma C.1. For the last two ones, differentiating totally the F.O.C. condition for the second period, we get:

\[
\left\{ E \left[ u(c_2) \frac{\partial e}{\partial e} \right] - \psi''(e_2) \right\} de - E \left[ u'(c_2) \frac{\partial e}{\partial e} \right] \{ Rdd + dR \} = 0
\]

The first term is negative (S.O.C) while the second is positive by lemma C.1. Hence the result.

**Proposition 4.3:** assume that \( \frac{\partial}{\partial \Delta w} < \frac{u''}{u} \). Then,

**Proposition C.3** (i) the effort level in the first period, for a given federal tax rate is smaller than without local borrowing;

(ii) the effort level in the second period, for a given federal tax rate, can be higher or lower depending on the debt level;

**Proof.** (i) we use a revealed preference argument. We show that:

\[
E \left[ u(c) \frac{\partial e}{\partial e} \mid e, w^* \right] \leq E \left[ u(c) \frac{\partial e}{\partial e} \mid e, w^*, d = 0 \right]
\]

Using lemma C.1, a sufficient condition is:

\[
\frac{\partial}{\partial w} \left( u \left( w \left( 1 - \frac{\tau}{2} \right) + \frac{\tau}{2} w^* + d(\Delta w) \right) - u \left( w \left( 1 - \frac{\tau}{2} \right) + \frac{\tau}{2} w^* \right) \right) \leq 0
\]

Or:

\[
u' \left( w \left( 1 - \frac{\tau}{2} \right) + \frac{\tau}{2} w^* + d(\Delta w) \right) d'(\Delta w) +
\{ u' \left( w \left( 1 - \frac{\tau}{2} \right) + \frac{\tau}{2} w^* + d(\Delta w) \right) - u' \left( w \left( 1 - \frac{\tau}{2} \right) + \frac{\tau}{2} w^* \right) \} \left( 1 - \frac{\tau}{2} \right) \leq 0
\]

The first term is always negative, as \( d'(\Delta w) < 0 \). The second one depends on the sign of \( \Delta w \).

If \( \Delta w > 0, \ d < 0 \) and the second term is positive. If \( \Delta w < 0, \ d > 0 \) and the second term is negative. We assume the first term always dominate. A sufficient condition is: \( \frac{\partial}{\partial \Delta w} < \frac{u''}{u} \). ■

202
Chapter 5

Consumption over the Lifecycle

Joint with Jonathan Parker

1 Introduction.

This chapter analyzes consumption and savings behavior of households over their working lives, focusing on estimation of age-specific consumption functions and their implications. We are motivated by four observations.

First, household consumption and savings decisions are arguably among the most important determinants of economic growth and business cycles. Consumer expenditures account for two thirds of national output and a large percent of output fluctuations. The difference between consumption and income—savings—determines the stock of wealth, which in turn determines the interest rate, the level and perhaps the growth rate of output. To understand business cycles and growth, we must thus first understand household consumption behavior.

Second, better methodology, data, and creative use of natural experiments have lead to more frequent and more convincing rejections of the most widely-used model of consumption behavior, the certainty equivalent life-cycle hypothesis (henceforth, CEQ
LCH). At the individual level, empirical studies typically test the central implication of the CEQ LCH -the Martingale hypothesis-, by testing whether consumption responds to expected changes in income. Despite generally poor-quality individual data on consumption, the CEQ LCH is often rejected.\(^1\) On aggregate data, the CEQ LCH is even more convincingly rejected.\(^2\)

Third, consumption smoothing does not seem to be a good characterization of low frequency consumption movements and savings behavior. According to the 1992 Survey of Consumer Finances (SCF), only 15% of the respondents reported retirement as their primary motive for saving while 42% cited liquidity needs.\(^3\) Life-cycle savings seem to occur late in the working lives of consumers. Median holdings of very liquid assets for households under age 50 are $3,900 while median holdings of non-housing non-business wealth are just under $13,000.\(^4\)

Lastly, recent theoretical work (Zeldes (1989), Deaton (1991), and Carroll (1992)) demonstrates that the Martingale hypothesis and consumption smoothing can be bad approximations to consumer behavior when agents face large amounts of individual uncertainty. Carroll (1993a) and Hubbard, Skinner and Zeldes (1994) among others have shown that income uncertainty can generate a positive covariance between expected income changes and consumption at low and high frequencies through precautionary savings.

In this chapter, we examine individual consumption data, estimate a structural model

\(^1\)It should also be noted that many of these papers technically test the permanent income hypothesis rather than the LCH. The differences can be pronounced at the aggregate level, but both theories predict no response of consumption to expected changes in income at the household level.


\(^3\)See Carroll (1993a) and Carroll (1992).

\(^4\)As reported in Carroll and Samwick (1994). This does not include Social Security and pension wealth, which constitute a large fraction of wealth at retirement. Note that distribution of wealth and its constituents is always strongly skewed, with a mean far exceeding the median.
with realistic levels of income uncertainty, and finally reinterpret life-cycle consumption and asset accumulation behavior within the context of the model. We measure, exploit and analyze the systematic age-pattern of consumption profiles. Age-heterogeneity in consumption behavior results from (a) the interaction between, and relative strengths of, retirement and precautionary motives for saving at different ages and (b) the changing slope of the income profile. We are successful along several dimensions.

First, we provide new evidence on the failure of the Certainty Equivalent (CEQ) LCH at the microeconomic level. We do this by demonstrating that consumption age-profiles averaged across time and households are not flat and are related to expected profiles in income. Importantly, this result still holds after controlling for family composition and cohort effects --- two potential reasons for the observed hump-shape of consumption. Our approach involves using the best available data on consumption expenditures in the US, the Consumer Expenditure Survey (CEX) from 1980 to 1993, giving us data on around 40,000 households. Using weak identifying assumptions, we construct consumption and income profiles across the working lives of "typical" men of five different educational attainments and five different occupational groupings. Consumption and income profiles are both significantly hump-shaped, and consumption tracks income reasonably well early in life.

Second, embedding realistic income uncertainty into the canonical model of life-cycle consumer behavior substantially improves the fit of the predicted life-cycle consumption profile. To demonstrate this, we write down a consumer maximization problem with a retirement period and explicit income uncertainty. The solution to this model will be the standard CEQ LCH consumption rules when uncertainty is ignored. Consumption will depend on the interest rate, the intertemporal elasticity of substitution, the discount factor and the present discounted value of income. However, under uncertainty, consumption will also depend on the path of expected income. Thus consumer behavior will vary systematically over the life cycle. When expected income growth and the discount
rate are low relative to the interest rate, consumers' behavior will remain similar to that of standard life-cycle consumers. If, on the other hand, expected income growth or the discount rate are large relative to the interest rate, consumers will behave as "buffer-stock" agents, consuming roughly their income and saving only small amounts to buffer against bad income draws. As households age, income growth declines. Consequently, the retirement savings motive will enter consumers' horizons. They will save more and behave more like certainty-equivalent consumers. The model can potentially deliver average consumption profiles which are more concave than income profiles. It is important to emphasize that we do not assume our results here. With a sufficiently low discount rate, the average consumption profile would be very similar to that of the certainty equivalent case.

Positing that the average income profile for a given group corresponds to the expected income profile and incorporating calibrated individual-specific income shocks, we estimate consumption functions for consumers in each occupation and education group. By simulating the lives of many consumers using these consumption functions we create predicted average consumption profiles. We then estimate the discount rate and the intertemporal elasticity of substitution by a Method of Simulated Moments procedure. The average household has an intertemporal elasticity of substitution of 2.04 and a discount rate of 3.9%. The discount factor is tightly estimated, and the estimated discount rates decline weakly with educational attainment. It is worth stressing that the estimated coefficients are within a "reasonable" range. In particular, buffer-stock behavior arises early in life due to the steepness of the income profiles at young ages.

The model fits the data quite well and does an excellent job of capturing the main features of the consumption profiles. To the best of our knowledge, this represents the first structural estimation of consumption functions over the life cycle which incorporates

---

5The latter estimate is very sensitive to the assumption about the real after tax interest rate, 3% in our benchmark analysis. We discuss robustness issues in section 5.
precautionary savings.

Third, we find strikingly different consumption functions for households at different ages: consumers behave like "buffer-stock" consumers early in their working lives and more like CEO LCH consumers as retirement nears. We show that households make the transition from buffer-stock to LCH behavior just after age 42. This confirms our initial intuition: relative movements of the consumption and income profiles reveal a great deal of information about the relative strength of the two savings motives. We conclude that a large fraction of consumers consists of target savers, for whom the Euler equation, as typically tested, should be expected to fail. This is, in part, a confirmation of Carroll (1993a) and (1993b) which argue, based on asset data, that buffer-stock models apply only to households before ages 45 to 50.

Fourth, this chapter contributes to the debate on the determinants of wealth accumulation. In our model, saving and consumption at each age are determined by the interactions between precautionary and retirement savings motives. Defining all wealth accumulation at retirement as life-cycle savings, we can decompose saving into precautionary and life-cycle saving. Early in life, households would like to borrow against expected future labor income. Consequently, life-cycle savings are negative. However, uncertainty causes households to build a buffer stock of savings, implying that precautionary saving is positive. Late in life, labor-income uncertainty is mostly resolved, and consumers run down their buffer stocks, while retirement saving becomes positive and large. Thus, the calibrated model matches the basic features of asset data. Both the fact that most households hold few, if any, assets and the fact that most households do not start saving for retirement until late in life have often been interpreted either as evidence against the LCH or evidence against forward-looking consumers. Our fitted model suggests that these facts are instead consistent with the LCH augmented to include income

---

6We do not enter the debate on the relative importance of retirement versus bequest savings. Implicitly, our retirement savings measure will include both.
uncertainty and consistent with forward-looking optimizing behavior.

Our chapter builds on many previous studies of life-cycle consumption behavior.

Several papers have used micro-consumption and income data to construct life-cycle profiles of consumption and income. Kotlikoff and Summers (1981) construct synthetic life-cycle profiles of consumption and income and present some evidence that consumption falls significantly below income only after age 50. Carroll and Summers (1991) report that consumption tracks income across countries, education and occupation groups, providing additional evidence that life-cycle savings do not seem to occur until late in life. Both studies find that consumption tracks income over the lives of households until around age 50. However, these studies are using cross-sectional data to infer time-series behavior. Thus, the close correlation between consumption and income may come from cohort-effects: on average, young families have larger lifetime resources and hence consume more.7 Further, changes in family size and consumer-needs over the life cycle may impart a hump-shape to consumption which does not come from a failure of CEQ consumption smoothing. We address these issues by adjusting both consumption and income for life-cycle changes in family size and for cohort effects.

More recently, Attanasio and Browning (1995) and Attanasio, Bank, Meghir and Weber (1995), using data from the UK Family Expenditure Survey (FES) and the US Consumer Expenditure Survey (CEX) respectively, have examined life-cycle behavior adjusted for cohort and family size effects. Attanasio et al. (1995) shows that the residuals from a regression of consumption on family composition and labor supply variables are uncorrelated with age. However, if the true consumption profile is hump shaped over the life cycle, this regression suffers from an omitted variable bias, which will incorrectly assign the hump to changes in demographics. Both papers also draw life-cycle implications from certainty equivalent Euler equation estimation under flexible representations.

of preferences. They do not reject the certainty equivalent life-cycle model.\textsuperscript{8} Here again, however, the instruments used in the Euler equation estimation are likely to be correlated with the omitted precautionary term. This overestimates the share of the consumption hump attributed to labor supply and family size variables.\textsuperscript{9}

We are also building on previous studies which parameterize and simulate life-cycle consumption models with uncertainty. Hubbard et al. (1994) and Carroll (1993a) show that the optimal consumption choices of consumers lead to profiles which are hump-shaped and track income over the early part of life for some parameterizations. Hubbard, Skinner and Zeldes (1995) go further and choose simulated profiles so as to try to reproduce constructed profiles of assets over the life-cycle.\textsuperscript{10} They re-interpret low-asset holding by most households as driven by means-tested government programs.\textsuperscript{11} Our approach goes beyond those studies by estimating a structural model of consumption.

Further, Palumbo (1994) uses individual consumption, income and asset data to estimate individual consumption functions for retirees. We choose to rely on average profiles precisely because we do not believe that the individual-level data are of sufficient quality to support the employed technique in general.\textsuperscript{12}

Finally, nearly all previous Euler equation estimations of these parameters ignore

\textsuperscript{8}Since adjusted profiles are not reported, it is not clear whether their finding is merely the result of adding covariates - labor supply and family size - so that the instrumental variable technique they employ no longer has the power to detect the relationship. We note that one of the best papers which uses occupation and education to predict expected income growth, Lusardi (1993), finds that consumption does track income.

\textsuperscript{9}Both of these papers separately deflate components of consumption, which can eliminate a consumption-income parallel. If liquidity constraints are binding or consumers are buffer-stock then \textit{nominal} consumption will track \textit{nominal} income. Finally, both papers look only at nondurable consumption, which, as we demonstrate and discuss, is not simply a scaled down version of total consumption.

\textsuperscript{10}These authors do not correct for family-size or cohort effects.

\textsuperscript{11}We do not address this alternate interpretation, but simply note that we believe that the heterogeneity in skills, abilities, and wealth across people starting their working lives makes the low-asset trap of their model very relevant for a small subset of the population and much less relevant for the typical household. See also Carroll and Samwick (1994) for a critique of the implications of their approach.

\textsuperscript{12}For instance, Palumbo (1994) must use a scaled-up measure of food consumption as his measure of household consumption and must make various assumptions about each individual's expected health dynamics.
the precautionary term in the Euler equation, a potentially serious flaw. Two recent papers immune to this problem are Carroll and Samwick (1994) and Barsky, Juster, Kimball and Shapiro (1995). The former, using asset data and a theoretical framework similar to ours, finds that the discount rate is poorly identified. The latter, using survey questions about preferences over lotteries and income paths, estimates an intertemporal elasticity of substitution and a discount rate both lower than what we estimate. We are exploiting lower frequency movements in the data than typical Euler-equation tests. High-frequency Euler equation tests might reject the CEQ PIH, while the CEQ PIH could still be a reasonably accurate model for low frequency analysis.

The structure of the chapter is as follows. In section 2, we lay out a model of consumer maximization and its implications for the construction of consumption and income profiles. We describe the numerical dynamic programming techniques used to solve the model and present characterizations of optimal behavior. The third section describes the data, discusses empirical issues involved in constructing our life-cycle profiles, and presents graphs of the profiles. Section 4 introduces the method of Simulated Moments methodology for estimating the model. Finally, we present the results of the estimation and conclude. Appendices contain more detailed descriptions of the CEX data, the numerical optimization, and econometric technique.

2 Consumption Behavior with Stochastic Income.

We begin by setting up a model of consumer behavior incorporating two saving motives: retirement and precautionary. The life-cycle saving motive results from the finite lifetime of individuals and from their retirement period. Income uncertainty at the individual level provides incentives for precautionary savings.

By nature, we are dealing with a non-stationary problem, as expected income follows a deterministic path and the permanence of the shocks depends upon the consumer's
age. The consumer's program must be solved recursively, keeping track of consumption rules at each age. But this results precisely from our initial intuition: the systematic age-pattern of consumption functions will reflect the interaction of the two saving motives and will translate into some definite life-cycle profile. Conversely, the age-pattern of the profiles we construct will allow us to identify life-cycle consumption functions.

2.1 The Canonical Model with Labor Income Uncertainty.

Our starting point is the basic discrete-time, life-cycle model of consumption behavior. Consumers live for $N$ periods and work for $T < N$, where both $T$ and $N$ are exogenous and fixed. In every period $t < T$, the consumer receives a stochastic income $Y_t$ and consumes $C_t$. There is only one asset in the economy, totally liquid and yielding a constant gross, after-tax, real interest rate $R$. Our unit of analysis is the household. We assume that preferences take the standard additively separable expected utility form, with a discount factor $\beta$:

$$
U = E \left[ \sum_{t=1}^{N} \beta^{t} u(C_t, Z_t) + \beta^{N+1} V_{N+1}(W_{N+1}, Z_{N+1}) \right],
$$

where $W_t$ represents total financial wealth and $Z_t$ is a vector of household characteristics (e.g. family size). $V_{N+1}$ represents the value to the consumer of any assets left at the time of death, capturing any bequest motive. The consumer maximizes (2.1) given an initial wealth level $W_1$, and the constraint that terminal wealth is non-negative $W_{N+1} \geq 0$. The dynamic budget constraint is:

$$
W_{t+1} = R \left( W_t + Y_t - C_t \right). \tag{2.2}
$$

We further assume that the felicity function $u(\cdot, \cdot)$ is of the Constant Relative Risk Aversion (CRRA) form, with intertemporal elasticity of substitution $1/\rho$, and multiplicatively
separable in $Z$:\textsuperscript{13}

$$u(C, Z) = v(Z) \frac{C^{1-\rho}}{1-\rho}.$$  

If income were certain, the solution to this program would be standard: the consumer would choose a consumption path such that:

$$\frac{C_{t+1}}{C_t} = \left(\beta R \frac{v(Z_{t+1})}{v(Z_t)}\right)^{\frac{1}{\rho}}. \quad (2.3)$$

With constant individual characteristics, (2.3) implies a constant growth rate of consumption. Consumption increases (respectively decreases) over time when the interest rate is larger (respectively smaller) than the discount rate. The growth rate of consumption (as opposed to its level) is independent of the income profile. The consumption level is then determined by the lifetime budget constraint and the terminal value function. The desire to smooth consumption over the entire lifetime will induce households to save for retirement and for bequest during their working lives.

When individual characteristics vary over the life cycle, the growth rate of consumption may change accordingly. For instance, if the marginal utility of consumption increases with family size, consumption will grow as family size increases, and decrease as children leave the household. These variations in individual characteristics may induce a positive correlation between consumption and income over the life cycle.

The certainty (or certainty-equivalent) LCH provides extremely valuable insights on the determinants of consumption and savings. However, by deliberately assuming away individual uncertainty, it may be missing an important part of consumer behavior.

With individual income uncertainty and prudence, households will hold precaution-

\textsuperscript{13}Equivalently, $\rho$ is the coefficient of relative risk aversion. This is a well known feature of additively separable expected utilities. This chapter will not explore alternatives such as non-expected utilities (Kreps and Porteus (1978)) or habit formation (Heaton (1990)).
ary savings to insure themselves against future contingencies. The variation in this precautionary motive has far-reaching and striking implications. The main consequence of income uncertainty is to increase the slope of the consumption profile (provided that consumers are prudent). Hubbard et al. (1994) demonstrate that this uncertainty can lead to hump-shaped consumption profiles as households save for precautionary reasons early in life and run down these assets during retirement due to lower levels of uncertainty and an increased probability of death. Carroll (1992) and Deaton (1991) analyze the case in which consumers are also *impatient*: in the absence of uncertainty, households would like to borrow in order to finance a high level of current consumption. In addition, Deaton (1991) imposes liquidity constraints while Carroll (1992) sets up a model in which consumers choose never to borrow. In either rendition, assets play the role of a buffer stock against bad income shocks. Consumers have a target level of liquid assets, above which impatience dominates and assets are decumulated, and below which the precautionary motive dominates and assets are accumulated. Thus the theory predicts a correlation of expected income growth and consumption growth at both high and low frequency. Over the life-cycle, consumption will appear to track income.

In the rest of the chapter, we explicitly incorporate uninsurable idiosyncratic income uncertainty in addition to our retirement saving motive. We adopt Carroll's (1992) formulation, and decompose the labor income process into a permanent, $P_{jt}$, and a transitory, $U_{jt}$, components (where $j$ indexes occupation and education groups):

$$Y_{jt} = P_{jt}U_{jt}$$  \hspace{1cm} (2.4)

$$P_{jt} = G_{jt} P_{jt-1} N_{jt}$$

The transitory shocks, $U_{jt}$, are independent and identically distributed, take the value 0 with probability $p \geq 0$, and are otherwise log-normally distributed so that $\ln U_{jt}$ has mean zero and variance $\sigma^2_{u_{jt}}$. The log of the permanent component of income, $\ln P_{jt}$,
evolves as a random walk with drift. $G_{jt}$ is a deterministic growth factor (specific to age $t$ and group $j$) while $\ln N_{jt}$, the shock to the permanent component of income, is independently and identically normally distributed with mean zero and variance $\sigma^2_{nj}$.\textsuperscript{14} Thus income evolves as a nonstationary, serially correlated process, with both permanent and transitory shocks, and a positive probability of zero income in every period.\textsuperscript{15}

Two points are worth noting. First, permanent shocks are only as permanent as the length of the working life: all shocks are ultimately transitory, as consumers retire and die. In the last working period, transitory and permanent shocks are equivalent. As a consequence, the propensity to consume out of “permanent” shocks to income will decrease with age, a point emphasized by Clarida (1991). This property holds true for the CEQ LCH also. Second, in this setup consumers will choose never to borrow against future labor income. This follows from (a) there being a strictly positive probability that labor income will be arbitrarily close to zero for the rest of the working life and (b) the Inada condition $\lim_{c \to 0} u'(c) = \infty$.\textsuperscript{16} It is important to note that this holds true, even when $p$, the probability of strictly zero income, is set to zero. Suppose the household were to borrow in the next to last working period. Then, with strictly positive probability it would be left without any wealth in the last working period. The household would then have an infinite expected marginal utility. Simple backward induction implies that it will never be optimal to borrow. Thus, in this setup, the precautionary motive acts as a self-imposed liquidity constraint.\textsuperscript{17}

Going from the model to the data, we need to make three assumptions. First, note that in order to solve the consumer’s problem as stated, we need to specify both the

\textsuperscript{14}According to this decomposition, the permanent level of income is that level that would obtain without transitory shocks, not the present discounted value of future income streams. This matches Friedman’s (1957) interpretation.

\textsuperscript{15}While Abowd and Card (1989) found that change in labor income was best characterized by an $MA(2)$ process, they also found little gain in moving from an $MA(1)$ to an $MA(2)$.

\textsuperscript{16}A condition always satisfied with isoelastic utility and positive risk aversion.

\textsuperscript{17}If instead we had assumed a strictly positive lower bound on income, consumer could borrow up to the present discounted value of certain future income.
nature of uncertainty during retirement and a bequest function. While there have been
good attempts at modelling consumer behavior after retirement,\(^\text{18}\) we feel that we know
too little about the form that uncertainty takes after retirement to use our methodology
and draw inferences from post-retirement behavior. Uncertainty arises from different
sources—medical expenses, the timing of death and asset returns. Inter-vivos bequests
are important. Although these sources of uncertainty are also present to some extent in
the last working years, labor income uncertainty seems to be the most important form
of uncertainty. Further, high quality information on household asset holdings together
with consumption and income are not available. Given that investment income, social
security, and pensions represent the main sources of income during retirement, it is
currently difficult to establish consumption patterns as a function of total wealth.

Even with a proper treatment of retirement issues, one would have to make a guess
about the bequest function. Therefore we decided instead to make use of Bellman’s optimality principle, and truncate our problem at retirement.\(^\text{19}\) Defining the value function
at time \(t\), \(V_t\), our problem becomes:

\[
V_t (X_t, P_t, Z_T) = \max_{c_t, \ldots, c_T} E_T \left[ \sum_{t=T}^{T-\tau} \beta^{t-\tau} v(Z_t) \frac{C_t^{1-\rho}}{1-\rho} + \beta^{T-\tau} V_T (X_T, P_T, Z_T) \right]
\]

\[s.t. \quad X_{t+1} = R (X_t - C_t) + Y_{t+1}, \quad (2.5)\]

where we define cash on hand \(X_{t+1}\) as total available financial resources at time \(t + 1\):

\[X_{t+1} = W_{t+1} + Y_{t+1} = R (X_t - C_t) + Y_{t+1}.
\]

\(^{18}\)See Hubbard et al. (1994) and Palumbo (1994).

\(^{19}\)Our approach does not emphasize the relative importance of retirement versus bequest wealth. The salvage value function accommodates a bequest motive. See Modigliani and Brumberg (1956) and the ensuing debate with Kotlikoff and Summers (1981) about the relative importance of life cycle and bequest savings.

215
Second, the model imposes a single vehicle for precautionary and retirement savings, since there is only one asset. In practice, much of retirement savings is accumulated in the form of illiquid assets, only available after retirement.\textsuperscript{20} This suggests that the relevant model of consumption behavior should incorporate an additional asset which is illiquid and accessible only after retirement. However, this would substantially complicate the problem by introducing another control variable (how much to save in liquid versus illiquid assets) and state variable (illiquid assets). In order to keep our estimation procedure feasible, we instead assume that consumers will receive at the age of retirement some accumulated illiquid wealth, proportional to their permanent income. This illiquid wealth accumulates exogenously and cannot be borrowed against. Effectively, this imposes a borrowing constraint $W_t \geq 0$. We denote accumulated illiquid wealth as $H_t$ and total financial wealth after retirement as $A_t = H_t + W_t$.

Lastly, we need to postulate a salvage value function which summarizes the consumer's problem at retirement time. We choose a functional form which maintains the tractability of the problem and is flexible enough to allow robustness checks:

$$V_T (X_T, H_T, Z_T) = k v (Z_T) (X_T + H_T)^{1-\rho}$$ \hfill (2.6)

This functional form is exactly correct if the only source of uncertainty after retirement is the time of death. When we move to estimation, we will calibrate the parameters of the associated consumption rule at retirement using information on consumer wealth, income, and consumption.

\textsuperscript{20}Social security wealth is definitely illiquid and is only available as annuities after retirement. Early withdrawal of pension and savings vehicles targeted for retirement purposes, such as IRA's, 401k plans and Keogh, is often penalized, if allowed at all. One might also consider housing wealth as part of retirement wealth. Empirical evidence suggests that households run down their housing wealth only extremely late in life.
2.2 Solving for Optimal Consumer Behavior.

The setup of the problem combined with our particular choice of retirement value function makes the problem homogeneous of degree $1 - \rho$ in the permanent component of income. Since a second state variable would render our estimation procedure unfeasible, we assume here that individual characteristics are constant throughout the life cycle: $Z_t = Z = 1$. Furthermore, since our data does not track households over time, we cannot calibrate the family-size process for each household over its life cycle. We will instead directly control for family effects when constructing our profiles. This allows us to write the optimal consumption rule as a function of a single state variable, $x_t$, the ratio of cash on hand to permanent income:

$$
x_{t+1} = x_{t+1}/P_{t+1} = (x_t - c_t) \frac{R}{G_{t+1}N_{t+1}} + U_{t+1}. \quad (2.7)
$$

We can then derive the Euler equation in any period prior to retirement:

$$
u'(c_t(x_t)) = \beta RE \left[ u'(c_{t+1}(x_{t+1})G_{t+1}N_{t+1}) \right] = \beta R \{ p E \left[ u'(c_{t+1}(x_{t+1})G_{t+1}N_{t+1}) \mid U_{t+1} = 0 \right] + (1 - p) E \left[ u'(c_{t+1}(x_{t+1})G_{t+1}N_{t+1}) \mid U_{t+1} > 0 \right] \}, \quad (2.8)
$$

where lowercase letters are normalized by the permanent component of income, and $c_t(x_t)$ represents the optimal consumption rule at time $t$ as a function of normalized cash on hand $x_t$. Next period expected marginal utility can be decomposed according to Bayes formula into expected marginal utility conditional on zero future income, and expected marginal utility conditional on a strictly positive income.

The solution to the consumer problem consists of a set of consumption rules $\{c_t(x_t)\}_{1\leq t\leq T}$. In the last working period, under our previous assumptions, consumption will be linear.
in cash on hand:
\[ c_T(x_T) = \gamma_0 + \gamma_1 x_T, \]  
(2.9)

where \( \gamma_0 = \gamma_1 h_T. \) Consumption in the next to last period is then found as the solution to (2.8) for all values of cash on hand, where we replace \( c_T \) using (2.9). Solving recursively generates \( c_{T-1}, \ldots, c_1. \) A complete description of the solution method is provided in Appendix A.

2.3 Characterization of Individual Consumption Behavior.

Figure 2.1.a shows the consumption rules at various ages, when the permanent income profile is flat \( (G_t = G = 1) \), there is no retirement period \( (\gamma_0 = 0, \gamma_1 = 1) \), working life starts at age 25 and ends at age 65, and consumers are impatient. When permanent income growth is constant, the finite horizon problem converges in the limit to the infinite horizon one, as we move further away from retirement. Consumption is always positive, increasing and concave in cash on hand. One can also show that cash on hand can only increase if the income draw is sufficiently large. Early in life, households will exhibit the standard buffer stock behavior: for low level of assets, typically less than the permanent component of their income \( (x \leq 1) \), households will consume most, but never all, of their

---

21 In the case of full certainty after retirement, it is straightforward to show that: \( \gamma_1 = \frac{1}{\lambda N + 1} \) where \( \lambda = \beta^1 = \beta R^{1/\rho - 1}. \)

22 The consumption rules have to be found numerically, as no closed form solution exists for this problem. We did this using a discretization method (see Judd (1993)). We solved the Euler equation (2.8) recursively on a grid of normalized cash-on-hand values. The consumption function was then interpolated between the points on the grid. In order to capture the curvature of the consumption rules at low values of cash on hand, the discretization grid was finer for \( x \in [0, 2] \). See Caballero (1990) for a closed form solution under exponential utility.

23 Other relevant parameters are \( \beta = 0.931, \rho = 1.13 \) and \( R = 1.05. \) These parameters generate buffer-stock behavior. Income uncertainty is the average amount, discussed later, and presented in Table 5.3.

24 In other words, the solution to the supremum problem in the infinite horizon case is obtained as the fixed point of the associated functional equation. Theorem 9.12 in Stokey, Lucas and Prescott (1989) applies even though returns are unbounded, as long as the discount factor is strictly less than 1.

25 This condition is analyzed in more details in Deaton (1991) and Ayagari (1993). In the infinite horizon case, this guarantees that cash on hand has an ergodic distribution.
financial wealth, and move to the next period with a very low level of assets. At high levels of cash on hand, households will consume a smaller fraction of cash on hand, but always enough so that they expect to run down their assets.

As death nears, the consumer faces less and less uncertainty from labor income shocks. It is then rational for an impatient consumer to start running down their buffer of assets: the consumption rules converge progressively towards the 45^0 line. In the last period, obviously, the household consumes everything. Consumers save only in order to buffer income shocks. We note also that from age 25 to 55, our consumer has roughly the same consumption rule. Buffer stock savings are run down quite late in life.

Figure 2.1.b shows consumption rules, for a household facing the same expected income growth, now assuming that the retirement consumption rule, at age 65 is more realistically characterized by:26

\[ \gamma_0 = 0.384, \quad \gamma_1 = 0.049. \]

The two households have the same consumption rules early in life—governed by the common solution to the infinite horizon problem. In other words, both households will behave as standard buffer stock households in their youth. Now however, the agent will have to accumulate enough wealth for retirement purposes. As retirement nears, savings must increase. Note also that with impatient consumers, the retirement savings motive matters only late in life.

Figures 2.2.a and 2.2.b display randomly drawn profiles of consumption for households facing typical paths of income, retirement rules, and income uncertainty.27 In both profiles, consumption tracks income early in life, and diverges later in life. Notice further that unexpected transitory shocks are better smoothed later in life, despite the fact that they contain greater information about total resources for the remainder of the life.

---

26 We discuss calibration of \( \gamma \) in section 4.
27 What is typical will be discussed in detail in Sections 4 and 5.
Smoothing is easier later in life since retirement saving also acts as a large buffer.

What are the necessary conditions to generate buffer-stock behavior early in life? Previous characterizations have addressed the problem in a stationary environment. Specifically, Deaton (1991) defines buffer-stock consumers as consumers who would borrow against future income, were it not for uncertainty. In the CEQ LCH, the slope of normalized consumption is given by (see (2.3)):

$$\ln \left( \frac{c_{t+1}}{c_t} \right) = \frac{1}{\rho} \ln (R \beta) - \ln (G) \approx \frac{1}{\rho} (\tau - \delta) - g,$$

where $G$ is the gross rate of income growth. The ratio of consumption to income increases whenever the right hand side is positive. A higher preference for the present or a lower interest rate will make buffer stock behavior more likely. However, households with a low discount rate and facing a high interest rate may still decide to behave as buffer-stock consumers, saving only a small fraction of their resources if their income profile is steep enough.

With i.i.d. income growth $G$, uncertainty, and an infinite horizon, Deaton (1991) shows that, agents are buffer-stock if and only if:

$$R \beta E \left[ G^{-\rho} \right] < 1.$$  

As Carroll (1992) emphasizes, buffer-stock agents have a desired level of cash on hand relative to permanent income. The existence of such target is both necessary and sufficient for buffer stock behavior in the infinite horizon framework. At low levels of assets, agents on average save to build the buffer, and cash on hand is expected to increase. For large levels of normalized cash on hand, the precautionary motive vanishes and agents increase consumption. This target level of cash on hand is defined as the fixed point of the mapping from current to expected future cash on hand.\(^{28}\)

\(^{28}\)In our finite horizon framework, this characterization will be necessary but not sufficient in the
\[ \bar{x}_t = E_t [x_{t+1}|x_t = \bar{x}_t]. \] (2.12)

Figure 2.3 presents the mapping \( \bar{x}_t \rightarrow E_t [x_{t+1}|x_t = \bar{x}_t] \) at various ages. It is convex, increasing and initially above the 45° line, so that in general it need not have a fixed point.

3 Data and Consumption and Income Profiles.

3.1 Profile Construction Methodology.

The construction of our profiles is motivated by the model presented in the previous section. We want the income profiles to be usable as inputs to the consumer optimization problem and the consumption profiles to be comparable to average consumption paths from the optimal program. That is, we are interested in constructing the income growth profile \( \{G_t\}_{t=25}^{65} \) and the average consumption profile \( \{\bar{C}_t\}_{t=25}^{65} \). Consumption at a certain age \( t \) will be the log-average across the distribution of cash-on-hand, permanent component of income, and consumer characteristics for that age:

\[ \bar{C}_t = \exp \left[ \int \int \ln c_t(x_t) P_t dF(x_t) dF(P_t) \right] \]

In order to construct such a profile, we must address three issues. First, due to its excessive noisiness, we do not exploit the limited panel nature of the Consumer Expenditure Survey, but instead rely on data from repeated cross-sections. Consequently, in our sample, birthyear and age will be correlated. Households observed at age sixty,
say, will have been born long before those we observe at young ages and will have on average lower lifetime resources, and lower levels of income and consumption at each age. Ignoring birthyear effects would lead to a negative bias in our estimate of the slope of income and consumption growth, especially, late in life.29 Second, the model refers to household consumption, adjusted for family size \((Z = 1)\). Since household size is hump-shaped over the life cycle, the correlation found in previous studies between consumption and income over the life cycle may disappear after correcting for family size.30 Finally, we are interested in exploiting some variation in expected income profiles across households. Thus, while we do conduct our exercise for the average household, we also focus our analysis on income and corresponding consumption profiles for subgroups of the US population defined by occupation and education groups. We assign households to these groups on the basis of the male head. Male labor force participation over the life cycle is high and stable, giving us more data and robust profiles.

We posit the following effects model for the natural logarithm of consumption for individual \(i\), in education/occupation group \(j\), of age \(a\), in year \(t\):

\[
\ln C_{ji}^{at} = f_i^a + \pi_j^a + b_i + y_i^t + \epsilon_{ji}^{at}\tag{3.13}
\]

That is, consumption of individual \(i\) is determined by a family size effect, \(f_i^a\); an effect, \(\pi_j^a\), specific to their age, \(a\), and education/occupation group, \(j\); a cohort effect, \(b_i\); a year effect, \(y_i^t\); and an idiosyncratic, individual effect \(\epsilon_{ji}^{at}\). We are interested in recovering

---

29 This point has been emphasized recently by Attanasio and Weber (1995).
30 Attanasio et al. (1995) make this argument, and find that, after correcting for family size, consumption is no longer significantly dependent on expected changes in income at high frequency. These authors do not demonstrate that the profiles are flat after making this correction. It is worth noting that both this paper and ours assume that family size is endogenous. If the buffer stock model is correct and having children is a form of consumption, then the decision to have children is affected by the expected path of income. By correcting for family size composition over the lifecycle, one is removing that portion of changes in consumption driven by expected income changes. As we will see, however, the profiles presented subsequently suggest that correcting for family size attenuates but does not eliminate the consumption income parallel.
By doing so, we create a profile which has a constant family size over the life cycle, and also correct for the fact that we do not actually follow the same individuals over their entire lives. As discussed in Deaton (1985), it is not possible to remove the linear component of the time and cohort effects without also removing the average (across education/occupation groups) age profile of consumption. We make the identifying restriction that time effects are related to business cycles and thus are well captured by the partial correlation of consumption with the regional unemployment rate.

Our procedure can then be summarized as follows. First we put all the data into real terms using the Gross Domestic Product implicit price deflator for personal consumption expenditures. Second we generate family size adjustments—\( f_i^a \) in equation (3.13)—and apply them to all the consumption and income data so that households have a constant effective size over the life cycle. Then, to construct unsmoothed profiles, we estimate the following model, over households with male heads aged 25 to 65, by weighted least squares with weights based on the CEX population weights:

\[
\ln X_i - \hat{f}_i = \pi a_i + \eta b_i + \nu U_t + \tau Ret_t + \varepsilon_i, \tag{3.14}
\]

where \( X \) is either consumption or income, \( a_i \) is a complete set of age dummies crossed with education or occupation group dummies, \( b \) is a complete set of cohort dummies (less one), \( U \) is the Census region unemployment rate in year \( t \), and \( Ret \) is a dummy for each group which is equal to 1 when the respondent is retired. Profiles are constructed by predicting \( \ln X_i \) for each age and grouping, setting the cohort and unemployment

---

31 This follows from the annoying identity that interview year less age equals birthyear.
32 Again, it is important not to use different deflators for different items within consumption or for income and consumption. This could break the relationship between cash on hand and consumption in nominal terms which is the relationship predicted by the buffer-stock theory.
33 We construct \( f_i^a \) by running equation (3.14) without separate age effects by subgroup and with family size dummies on the right hand side. The effective family size is then 2.8 (the sample average). We also experimented with exogenous family size adjustments—assuming \( f_i^a \) is simply family size raised to the power \(-0.7\). This led to profiles which were noisier and flatter early in life.
rates are at their average values and the retirement dummies to zero. Smooth profiles are estimated by replacing the age and cohort dummies by fifth order polynomials, and extending the highest age to 70 to avoid some of the endpoint problems commonly encountered with polynomial smoothing.

Income profiles are used to construct estimates of \( \{G_t\} \) for the consumer problem. Recalling that, \( \ln Y_t^i = \ln P_t^i + \ln U_t^i = \ln G_t + \ln P_{t-1}^i + \ln N_t^i + \ln U_t^i \), after removing the cohort, family, and time effects, our procedure is in effect taking a sample average over a large number of individuals, \( M \), with the same characteristics:

\[
\frac{1}{M} \sum_{i=1}^{M} \ln Y_t^i = g_t + \frac{1}{M} \sum_{i=1}^{M} \ln Y_{t-1}^i + \frac{1}{M} \sum_{i=1}^{M} \ln N_t^i + \frac{1}{M} \sum_{i=1}^{M} \ln U_t^i - \frac{1}{M} \sum_{i=1}^{M} \ln U_{t-1}^i
\]

Applying the Law of Large Numbers, the probability limits of the last three term are all zero. Hence, we get

\[
\text{plim} \left( \frac{1}{M} \sum_{i=1}^{M} \ln Y_t^i - \frac{1}{M} \sum_{i=1}^{M} \ln Y_{t-1}^i \right) = \ln G_t.
\]

Thus simply first differencing our log-average income levels gives the income growth rates which are input into the simulated model. \( \hat{Y}_t = \exp\left[ \frac{1}{M} \sum_{i=1}^{M} \ln Y_t^i \right] \) provides an estimate of average income by age.

3.2 The Consumer Expenditure Survey and Our Use of it.

The main data source for the consumption and income profiles is the Consumer Expenditure Survey (CEX). The CEX contains information about consumption expenditures, demographics, income and assets, for a large sample of the US population. The Survey is conducted by the Bureau of Labor Statistics in order to construct baskets of goods for use in the bases for the Consumer Price Index, and has been run continuously since 1980. We use data from 1980 to 1993 from the family, member, and detailed expendi-
tory files. The survey is known to have excellent coverage of consumption expenditures, to have reasonable data on liquid assets, and to have income information of moderate quality. The survey interviews about 5500 households each quarter. In a household's first interview, the CEX procedures are explained to them and information is collected so that they can be assigned a population weight. They are then interviewed four more times (once every three months) about detailed consumption expenditures over the previous three months. In interviews two and five, income information is collected, and in the final interview asset information is collected. Families rotate through the process, so that about 25% of households leave and are replaced in each quarter. About half of all households make it through all the interviews.

Each household contributes one datapoint to our sample. For each household we construct a measure of household income and consumption, and assign it to an occupation group, an education group, a birth cohort, an interview year, and a Census region. In order to obtain a high quality sample which tracks men and has the required information, we drop a significant portion of the data and make a series of adjustments. A detailed description of the data preparation is contained in Appendix B, however, we will make note of three major points here, and then turn to our definitions of consumption and income.

First, we dropped households which are classified as incomplete income reporters, which had any of the crucial variables missing, or which reported changes in age over the course of the survey greater than one year or negative. Households dropped when constructing profiles by occupation remain in the education profiles. We did not use the occupational classifications of Armed forces, Service workers, and Farming, forestry, or fishing due to small cell sizes. Similarly we do not analyze the group of households with male heads holding less than 9 years of schooling due to very few younger households. Second, we dropped all households with male heads younger than 25 or older than 70,

Table 5.1: Cell Sizes for Consumption Profiles

<table>
<thead>
<tr>
<th></th>
<th>Households</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>43031</td>
<td></td>
</tr>
<tr>
<td><strong>Education Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some Highschool</td>
<td>4409</td>
<td>10.2</td>
</tr>
<tr>
<td>Highschool Degree</td>
<td>12906</td>
<td>30.0</td>
</tr>
<tr>
<td>Some College</td>
<td>10027</td>
<td>23.3</td>
</tr>
<tr>
<td>College Degree</td>
<td>6570</td>
<td>15.3</td>
</tr>
<tr>
<td>Grad/Prof School</td>
<td>6087</td>
<td>14.1</td>
</tr>
<tr>
<td><strong>Occupation Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial and Prof.</td>
<td>13215</td>
<td>34.3</td>
</tr>
<tr>
<td>Technical, Sales, and</td>
<td>6976</td>
<td>18.1</td>
</tr>
<tr>
<td>Admin. Support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision Production,</td>
<td>5061</td>
<td>13.1</td>
</tr>
<tr>
<td>Craft and Repair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operators, Fabricators,</td>
<td>7116</td>
<td>18.5</td>
</tr>
<tr>
<td>and Laborers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Employed</td>
<td>3479</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Note: this table shows counts used in constructing unsmoothed consumption profiles. More observations are used for smoothed profiles and fewer are available for income profiles.

since as discussed above we are choosing to focus on the working life. Cell sizes are reported in Table 5.1. Third, while topcoding is very infrequent in consumption information, the household annual income variable reflects summation over a topcoded item for roughly half a percent of our households. Since, in most years, topcoding occurs at $100,000 in income subcategories, reported individual annual labor income is the source of almost all income topcoding problems. However, households are also asked the gross amount of their paycheck and what length of time-period this paycheck covers. By multiplying these two variables together, we construct a second measure of annual labor income. Topcoding on this variable occurs only for a few cases. We correct our measure of after-tax family income by replacing the reported annual labor income in family income with our constructed measure whenever the family income variable is topcoded. We are able to correct almost all topcoding.

Finally, we consider how best to construct measures of income and consumption which match the concepts in the theoretical model. First, we choose to define consumption
as total household expenditures less those on education, medical care, and mortgage interest. These categories of expenditure do not provide current utility but rather are either investments or negative income shocks.\(^{35}\) They are also excluded from our income definition.

It should be noted that our model refers to nondurable consumption at annual frequencies. Since we are averaging expenditures across a large number of individuals and looking across one-year horizons, the distinction between durables and nondurables is less likely to matter. Since the buffer stock model gives strong predictions about consumption tracking income, it is important not to break the consumption-income link when studying consumption.\(^{36}\)

Our measure of income is comprised of after-tax family income less Social Security tax payments, mortgage interest, expenditures on medical care, spending on education, pension contributions and after-tax asset and interest income. For the first five adjustments, the related expenditures do not provide current utility but are either non-liquid investments or, in the case of medical care, simply losses of income. Further these expenditures involve a large amount of commitment and are hard to substitute intertemporally. Were we to include pension contributions in income, for example, our measure of liquid assets which could be used to buffer bad income shocks would include pension wealth. We remove asset income since the input to our theoretical model is a profile of income net of liquid asset returns. Attanasio (1994) looks at savings behavior and finds large savings by the typical household over the entire life. If we consider a larger measure of income and take the relative levels of consumption and income seriously, we also find significant savings over the life cycle. Our preferred interpretation of pension and housing wealth however is that it is illiquid, and thus that the typical household has little in liquid assets.

\(^{35}\) We are arguing here that user cost of housing -repairs, maintenance, utilities, and housing services-captures the expenditures made for consumption on housing.

\(^{36}\) When computing power has increased, this issue is likely to be surmountable by adding a state variable for the stock of durable goods and taking a stand on the size and relevance of adjustment costs.
with which to buffer income shocks early in life.

3.3 Life Cycle Profiles.

Figure 3.1 presents the estimated consumption and income profiles for our entire sample.\textsuperscript{37} Even after correcting for cohort, time, and family effects, both profiles are still hump shaped and still track each other early in life. Consumption lies above income from age 25 to 28. Reflecting on our own experiences, we may interpret this as underreporting the assistance which is provided by intergenerational transfers early in life. After these first few years, consumption rises with income from age 30 to age 45, when consumption drops significantly below income. This tracking is however a lot less than is observed in profiles constructed by simply averaging cross-sections. As stated above, the two main reasons for this are the changes in family size over the life cycle and the different wealth and incomes of different cohorts. Figure 3.2 displays the profile of average family size over the life cycle. Figures 3.3a and 3.3b present the consumption and income profiles without the cohort adjustment and without the family size adjustment, respectively. In each case the unadjusted profiles are more hump shaped, and seem to track each other more closely. These profiles are more directly comparable to those shown in Kotlikoff and Summers (1981) and Carroll and Summers (1991) which do not correct for family size over the life cycle or for cohort effects.

Despite the fact that the asset data in the CEX is of lesser quality than data from sources like the Survey of Consumer Finances (SCF), we can see an interesting age pattern in the profile of asset accumulation, which mimics that found in more accurate surveys. Figure 3.4 shows a life-cycle profile of the ratio of total liquid assets to income. Liquid assets in the CEX are the value of holdings of stocks and bonds, and the cash

\textsuperscript{37}We get reasonable relative levels of consumption and income as does Attanasio (1994), who uses relative levels in his analysis of saving in the US. However, the levels may not tightly identified by the data. In our results section, we will present evidence on estimation which instead uses only information from the changes in income and consumption.
Table 5.2: F-Tests for Flatness of Consumption Profiles

<table>
<thead>
<tr>
<th>Whole Sample</th>
<th>F-Stat.</th>
<th>Education Group</th>
<th>Occupation Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some Highschool</td>
<td>637</td>
<td>Managerial and Professional Specialty</td>
<td>934</td>
</tr>
<tr>
<td>Highschool degree</td>
<td>1015</td>
<td>Technical, Sales, and Admin. Support</td>
<td>803</td>
</tr>
<tr>
<td>Some College</td>
<td>1023</td>
<td>Precision Production, Craft and Repair</td>
<td>676</td>
</tr>
<tr>
<td>College Degree</td>
<td>804</td>
<td>Operators, Fabricators, and Laborers</td>
<td>790</td>
</tr>
<tr>
<td>Grad/Prof School</td>
<td>533</td>
<td>Self-Employed</td>
<td>604</td>
</tr>
</tbody>
</table>

Note: the F-statistic is distributed as a $F(39, 43000)$. The critical value is 1.40.

held in savings and checking accounts. We observe that the typical household accelerates its building of a stock of liquid assets around age 45. This age pattern also shows up in the profiles of income and consumption and is the crucial feature that will help us to pin down the time-heterogeneity in consumer behavior: until around age 45, people consume roughly their income, saving very little in the form of liquid assets.

Figures 3.5 and 3.6 give some evidence that consumption and income track each other across subgroups of the population defined by education and occupation levels. These graphs are unfortunately noisy. However, despite the noise in the data, one can see that the occupation and education groups with the most pronounced humps in income present the most pronounced humps in consumption. Further, we can formally reject the null hypothesis that the consumption profiles are flat. Table 5.2 presents F-tests of the equality of 40 age dummies in a profile regression which also includes age (when appropriate, interacted with subgroup). We get very strong rejection of a constant slope in all our profiles.\textsuperscript{38} We could also proceed to test whether the income profiles are

\textsuperscript{38}Of course it is possible that we have omitted a key preference shifter which varies over the lifecycle. As Attanasio et al. (1995) note, labor supply is an obvious candidate. Thus, we also test for flatness conditioning on labor income variables. We include in the regressions, annual hours worked by the male head and annual hours worked by the female head. We are still able to reject flatness in the total and in all subgroups. We do not use profiles constructed after removing this correlation in our later analysis, since our theoretical model implicitly assumes that utility is additively separable in leisure and consumption expenditures.
significant in predicting consumption profiles across occupation and education groups. But this is essentially a Hall and Mishkin (1982)-style test of the CEQ consumption Euler equation. Lusardi (1993) performs such a test, merging Panel Study of Income Dynamics (PSID) data with the CEX consumption to get the best possible combination of data, and she rejects the null hypothesis that the consumption profiles are unrelated to the income profiles.

Finally, Figure 3.6 displays the profile for total, nondurable, and food consumption, all rescaled to the same mean. This figure demonstrates that total consumption over the life cycle is not simply a scaled up version of nondurable consumption. Expenditures on durable goods involve spending income when the good is purchased, and receiving a utility flow over time. To the extent that consumers would like to borrow against future income, purchases of durable goods tighten this constraint by moving expenditures forward in time relative to utility flow. If consumers are buffer stock when young and if nondurable and durable consumption are substitutes, then one would expect to see total consumption rising slower early in life and peaking later than nondurable consumption. The relative profiles of these consumption categories are at least consistent with what we shall conclude subsequently: that households are behaving as buffer stock consumers early in their lives.

4 Estimation Strategy.

4.1 Method of Simulated Moments (MSM) Estimation.

According to section 2, consumption at age $t$ for individual $i$ depends on cash on hand $x_i^t$, the realization of permanent component of income $P_i^t$, the entire path of expected permanent income growth $\sigma_T = \{G_t\}_{t=1}^T$, and the parameters of the consumption prob-
lem \( \bar{\theta} = (\beta, \rho, R, \gamma_0, \gamma_1, p, \sigma_u^2, \sigma_n^2) \), an 8x1 vector.\(^{39}\) In practice, it will not be possible to estimate directly all the elements of \( \tilde{\theta} \). Instead, we will calibrate most of the parameters using existing micro data and will focus on the estimation of the structural parameters of the utility function \( \beta \) and \( \rho \). In other words, we rewrite \( \bar{\theta} = (\theta', \chi')' \) where \( \theta = (\beta, \rho)' \) and \( \chi = (R, \gamma_0, \gamma_1, p, \sigma_u^2, \sigma_n^2)' \). The elements of \( \chi \) will be calibrated from micro data and we will estimate only the elements of \( \theta \). Defining the vector of state variable \( z^i_t = (x^i_t, P^i_t) \), for individual \( i \), we postulate the following data-generating process:

\[
\ln C^t_i = \ln C_t \left( z^i_t, \theta, \chi; \varnothing_T \right) + \epsilon^i_t = \ln \left( c_t \left( x^i_t, \theta, \chi, \varnothing_T \right) P^i_t \right) + \epsilon^i_t, \tag{4.15}
\]

where \( \epsilon^i_t \) is an idiosyncratic shock that represents measurement error in consumption levels and satisfies \( E \left[ \epsilon^i_t | z^i_t \right] = 0 \).\(^{40}\) We are interested in estimating \( \theta \). Direct estimation of (4.15) is not possible since we have only poor information on individual assets \( x^i_t \).\(^{41}\) We do however observe \( \chi \) and average consumption at each age, \( \ln \bar{C}_t \), as defined in the previous section. This suggests that we can look directly at the unconditional expectation of log-consumption at each age:

\[
\ln \bar{C}_t(\theta; \chi, \varnothing_T) \equiv E \left[ \ln C_t (z, \theta; \chi, \varnothing_T) \right] = \int \ln C_t(z, \theta; \chi, \varnothing_T) dF_t(z). \tag{4.16}
\]

This says that average consumption of households equals the average level of consumption over both values of cash-on-hand and levels of permanent income. Our approach consists in matching the \( T \) moment conditions:

---

\(^{39}\)For ease of notation and consistency with our theoretical model, we assume in this subsection that age runs from 1 to \( T \).

\(^{40}\)\( \epsilon^i_t \) may also encompass missing variables such as class. Our approach remains correct as long as \( \epsilon^i_t \) and the state variables are independent.

\(^{41}\)This is the approach taken by Palumbo (1994) who uses Maximum Likelihood and PSID data to estimate structural parameters during the retirement period. One could also estimate consumption functions non-parametrically, much like Gross (1994), using Kernel methods, estimates investment functions of firms facing liquidity constraints.
\[ E \left[ \ln C_t - \ln \hat{C}_t (\theta; \chi, \Theta_T) \right] = 0. \]

Defining the sample moment at age \( t \), \( g_t (\theta; \chi, \Theta_T) = \ln \hat{C}_t - \ln \hat{C}_t (\theta; \chi, \Theta_T) \), and \( g (\theta; \chi, \Theta_T) \) the vector of moments, the estimation procedure minimizes

\[ g (\theta; \chi, \Theta_T)'W_T g (\theta; \chi, \Theta_T) \]

where \( W_T \) is a weighting matrix. With a weighting matrix equal to the identity, this is equivalent to minimizing the distance between the average and the estimated profile. That is, considering the life-cycle profile, we minimize:

\[ S (\theta; \chi, \Theta_T) = \sum_{t=1}^{T} \left( \ln \hat{C}_t - \ln \hat{C}_t (\theta; \chi, \Theta_T) \right)^2. \quad (4.17) \]

Unfortunately we do not directly observe \( \ln \hat{C}_t (\theta; \chi, \Theta_T) \), since we do not observe the distributions of permanent income or of cash-on-hand. Further, we do not have analytic solutions for the consumption functions or how they change with alteration of our key parameters. Thus, instead of computing the actual expectation with respect to the true distribution of \( z \), we use a Monte-Carlo integration method and perform a Method of Simulated Moments (MSM) estimation. We draw a random sample of shocks to income \( \{U_i, N_i\}_{i=1}^{L} \), as defined in (2.4), calculate the associated paths of consumption and cash on hand, and compute the simulated sample average:

\[ \ln \hat{C}_t (\theta; \chi, \Theta_T) = \frac{1}{L} \sum_{j=1}^{L} \ln C_t (z_t^j, \theta; \chi, \Theta_T). \]
We then choose parameters to minimize:

\[ \hat{S}(\theta; \chi, \Theta_T) = \sum_{t=1}^{T} \left( \ln \hat{C}_t - \ln \hat{C}_t(\theta; \chi, \Theta_T) \right)^2. \]

In practice, we simulate \( \ln \hat{C}_t(\theta; \chi, \Theta_T) \) by running 20,000 independent income processes (temporary and permanent) for 40 years, and computing in each year the associated consumptions. That is, we average over 20,000 profiles as in Figures 2.2.a and 2.2.b.

Under the regularity conditions stated in Pakes and Pollard (1989) and Duffie and Singleton (1993), the MSM estimator \( \hat{\theta} \) is both consistent and asymptotically normally distributed. Denoting the number of observations at age \( t \) as \( I(t) \), \( I = \frac{1}{T} \sum_{t=1}^{T} I(t) \), \( \bar{g}_t = \ln \hat{C}_t - \ln \hat{C}_t(\theta; \chi, \Theta_T) \), and \( \bar{g} = (\bar{g}_1, ..., \bar{g}_T)' \):

\[ \sqrt{I}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, V), \]

where \( V \) is estimated by:

\[ \hat{V} = \left( \hat{D}' \hat{D} \right)^{-1} \hat{D}' \hat{\Omega} \hat{D} \left( \hat{D}' \hat{D} \right)^{-1} \tag{4.18} \]
\[ \hat{D} = \frac{\partial \bar{g}}{\partial \theta' \theta = \hat{\theta}}. \tag{4.19} \]
\[ \hat{\Omega} = \text{avar}(\bar{g}) \tag{4.20} \]

Efficient estimation is obtained when \( \hat{W}_T = \hat{\Omega}^{-1} \). In practice, we use \( \hat{W}_T = \hat{\Omega}^{-1} \).

This methodology also provides a useful overidentifying restriction test. If the model is correctly specified, the statistic

\[ \chi_{T-2} = I \bar{g}(\theta; \chi, \Theta_T)' \hat{\Omega}^{-1} \bar{g}(\theta; \chi, \Theta_T) \]

233
is distributed asymptotically as Chi-squared with $T - 2$ degrees of freedom.

As discussed previously, the levels of our profiles might be misestimated. To test the robustness of our results, we also estimate without using information on the level of consumption. Assuming that the bias is constant through time, the moment conditions become:

$$E \left[ \ln C_t - \ln \hat{C}_t (\theta; \chi, \Theta_T) - \alpha \right] = 0 \quad ; \quad \forall 1 \leq t \leq T,$$

for some unknown constant $\alpha$. Since we are not interested in estimating this parameter, we instead rewrite the moment condition in first difference:

$$E \left[ \Delta \ln C_t - \Delta \ln \hat{C}_t (\theta; \chi, \Theta_T) \right] = 0 \quad ; \quad \forall 2 \leq t \leq T,$$

where $\Delta$ is the time-difference operator. This amounts to performing the estimation in first differences, that is, minimizing:\[42\]

$$\tilde{S} (\theta; \chi, \Theta_T) = \sum_{t=2}^{T} \left( \Delta \ln \bar{C}_t - \Delta \ln \hat{C}_t (\theta; \chi, \Theta_T) \right)^2.$$

The rest of the procedure is identical, with the exception that we now have only $T - 1$ moments.

### 4.2 Remaining Calibration and Data.

In order to find the consumption rules for the consumers in our dynamic program, we must still specify the elements of $\chi$. Computing power currently limits us to searching over only two parameters, and to checking robustness and relative explanatory power as we change other parameters. We now present our calibration of the remaining parameters in our model.

[42 with an identity weighting matrix.]
1. The after-tax real interest rate is set as \( r = R - 1 = 3\% \) per annum. This is roughly the average real interest rate on high grade municipal bonds over the sample period of our data.

2. The variance of the permanent and transitory components of shocks to income, \( \sigma_u^2 \) and \( \sigma_n^2 \), are taken from Carroll and Samwick (1994). Carroll and Samwick (1994) estimate these parameters from the Panel Study of Income Dynamics (PSID), which provides repeated high-quality measures of household income. The estimation procedure is based on income differences of different lengths and correctly assigns the relative importance of transitory and permanent income shocks even in the presence of significant moving average correlation of transitory shocks, up to an \( MA(2) \). The procedure and data employed are designed to estimate the parameters for the income process in Carroll (1992)—that is, exactly the income process we have specified.\(^{43}\) Table 5.3 displays the variances of the permanent and transitory shocks across education and education groups.

3. We follow Carroll (1992) and set the probability of zero income to \( p = 0.005 \). This calibration comes from the PSID and again is estimated with the goal of calibrating exactly the income process which we are considering.

4. Given our assumption on the retirement value function, the optimal consumption function in the last working period is linear in the permanent component of income and the level of cash on hand, or in normalized terms: \( c_T = \gamma_0 + \gamma_1 x_T \). This results, as discussed earlier, from the assumption that post-retirement income—pension and Social Security income—is illiquid, cannot be borrowed against before retirement, and that the only source of uncertainty after retirement is the time of death. In order to calibrate the parameters of this consumption rule, we first construct an

\(^{43}\)The definitions of occupation in the PSID and CEX do not exactly overlap, so that we are required to make rather crude adjustments to one cell.
Table 5.3: Variance of Income Shocks

<table>
<thead>
<tr>
<th>Group</th>
<th>Variance of Permanent Shock</th>
<th>Variance of Transitory Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>0.0217</td>
<td>0.0440</td>
</tr>
<tr>
<td>OCCUPATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial and Prof. Speciality</td>
<td>0.0180</td>
<td>0.0357</td>
</tr>
<tr>
<td>Tech., Sales, and Admin. Support</td>
<td>0.0235</td>
<td>0.0361</td>
</tr>
<tr>
<td>Precision Prod., Craft, and Repair</td>
<td>0.0175</td>
<td>0.0432</td>
</tr>
<tr>
<td>Operators and Laborers</td>
<td>0.0299</td>
<td>0.0458</td>
</tr>
<tr>
<td>Self Employed</td>
<td>0.0165</td>
<td>0.0926</td>
</tr>
<tr>
<td>EDUCATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some Highschool</td>
<td>0.0214</td>
<td>0.0658</td>
</tr>
<tr>
<td>Highschool Degree</td>
<td>0.0277</td>
<td>0.0431</td>
</tr>
<tr>
<td>Some College</td>
<td>0.0238</td>
<td>0.0342</td>
</tr>
<tr>
<td>College Graduate</td>
<td>0.0146</td>
<td>0.0385</td>
</tr>
<tr>
<td>Graduate School</td>
<td>0.0115</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Source: Carroll and Samwick (1994).

income profile which adds Social Security and pension contributions to the income measure we consider. We do not use housing wealth to calibrate this parameter since most elderly do not run down the asset value of their housing. The difference between this profile and the consumption profile is accumulated at the assumed interest rate and gives an estimate of total resources at retirement, $W_T + H_T$. A similar calculation using our main income measure gives a measure of liquid assets at retirement, $W_T$. Finally we calculate the required parameters using date from smoothed profiles at retirement as:

$$\gamma_1 = \frac{C_T}{W_T + H_T + Y_T}$$

$$\gamma_0 = \gamma_1 \frac{H_T}{Y_T}.$$
### Table 5.4: Data for Retirement Consumption Rule

<table>
<thead>
<tr>
<th></th>
<th>Cons.</th>
<th>Income</th>
<th>Liquid</th>
<th>Total Assets</th>
<th>Total Resources</th>
<th>γ₀</th>
<th>γ₁</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOTAL AVERAGE</strong></td>
<td>18349</td>
<td>23839</td>
<td>164616</td>
<td>352490</td>
<td></td>
<td>0.384</td>
<td>0.049</td>
</tr>
<tr>
<td><strong>EDUCATION GROUP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some Highschool</td>
<td>13631</td>
<td>16112</td>
<td>48296</td>
<td>156788</td>
<td></td>
<td>0.531</td>
<td>0.079</td>
</tr>
<tr>
<td>Highschool Graduate</td>
<td>14556</td>
<td>17791</td>
<td>146852</td>
<td>311519</td>
<td></td>
<td>0.409</td>
<td>0.044</td>
</tr>
<tr>
<td>Some College</td>
<td>18768</td>
<td>24865</td>
<td>215380</td>
<td>427021</td>
<td></td>
<td>0.354</td>
<td>0.042</td>
</tr>
<tr>
<td>College Graduate</td>
<td>19628</td>
<td>27375</td>
<td>344287</td>
<td>607399</td>
<td></td>
<td>0.297</td>
<td>0.031</td>
</tr>
<tr>
<td>Graduate School</td>
<td>24410</td>
<td>32690</td>
<td>345442</td>
<td>665461</td>
<td></td>
<td>0.342</td>
<td>0.035</td>
</tr>
<tr>
<td><strong>OCCUPATION GROUP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial and Prof.</td>
<td>22461</td>
<td>37732</td>
<td>406053</td>
<td>706256</td>
<td></td>
<td>0.240</td>
<td>0.030</td>
</tr>
<tr>
<td>Tech., Sales, Admin.</td>
<td>21394</td>
<td>30809</td>
<td>288376</td>
<td>520212</td>
<td></td>
<td>0.292</td>
<td>0.039</td>
</tr>
<tr>
<td>Precision Prod., Craft</td>
<td>17010</td>
<td>26726</td>
<td>252064</td>
<td>436494</td>
<td></td>
<td>0.253</td>
<td>0.037</td>
</tr>
<tr>
<td>Operators, Laborers</td>
<td>15557</td>
<td>23438</td>
<td>229031</td>
<td>370511</td>
<td></td>
<td>0.238</td>
<td>0.039</td>
</tr>
<tr>
<td>Self Employed</td>
<td>19654</td>
<td>25350</td>
<td>9097</td>
<td>201430</td>
<td></td>
<td>0.658</td>
<td>0.087</td>
</tr>
</tbody>
</table>


The results are summarized in the Table 5.4.44

5. Since households generally begin life with some assets, we capture this by assuming that initial cash on hand, $x₀$, is equal to 0.3 times permanent income at age 25, a number consistent with the CEX.45

## 5 Estimation Results.

We first estimate both the discount factor and the intertemporal elasticity of substitution for the average household. We then turn to disaggregated results, by education and occupation groups. Lastly, we discuss the robustness of our results to the calibrated parameters.

---

44 We find significantly larger assets on both counts than appears in wealth data (Venti and Wise (1993)).

45 As it turns out, the results are mostly insensitive to this assumption. See section 5.3.
5.1 Results for the Entire Sample.

We start by asking what the standard Life-Cycle theory would predict, assuming away all uncertainty. Although this constitutes a crude attempt at matching the data, it serves as a useful benchmark against which to evaluate the rest of our results. To give the best chance to the CEQ LCH, we perform first difference estimation of the CEQ LCH, not asking it to fit the mean of the consumption profile, as discussed at the end of section 4.1. Under certainty, equation (2.3) holds, implying, after controlling for individual characteristics, a constant growth rate of consumption over the working period:

$$\Delta \ln \bar{C}_t = \frac{1}{\rho} \ln (\beta R) \equiv \xi.$$  \hspace{1cm} (5.22)

We estimate $\xi$ from the coefficient on age in a least-squares regression on individual data. This procedure seems trivial only because of our earlier efforts to remove changing family-size and cohort effects. It is precisely this simplicity which gives the CEQ LCH its power. From our estimate of $\xi$, we use the delta method to recover the discount factor and its standard error, postulating a real interest rate of three percent and a coefficient of relative risk aversion of 0.49.\textsuperscript{46} The latter choice matters little since consumption is estimated to be nearly flat. The former matters a lot, and changes our estimates one for one.\textsuperscript{47} We estimate a discount rate of 2.56% with a standard error of 0.05.\textsuperscript{48} However, and not surprisingly given Figure 3.1, the certainty model performs poorly when it comes to explaining the dynamics of consumption across the life cycle. The estimated profile does not capture the hump shape in consumption, as Figure 5.1 demonstrates. Were

\footnotesize
\textsuperscript{46}This is the estimate from our MSM procedure with uncertainty. While in our stochastic model we are able to estimate the coefficient of relative risk aversion and the discount factor, here they are not separately identified.

\textsuperscript{47}This estimation procedure only captures the substitution effect, as is clear from (5.22). Income and wealth effects change the level of the consumption path, not its slope.

\textsuperscript{48}The standard error is not robust to serial correlation and probably is underestimated by an order of magnitude.

238
Table 5.5: STRUCTURAL ESTIMATION

<table>
<thead>
<tr>
<th>METHOD</th>
<th>LEVEL</th>
<th>LEVEL</th>
<th>DIFFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPTIMAL WEIGHTING</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9615</td>
<td>0.9625</td>
<td>0.9603</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.0043)</td>
<td>(0.0042)</td>
<td>(0.0048)</td>
</tr>
<tr>
<td>$\delta$ (%)</td>
<td>4.0017</td>
<td>3.8891</td>
<td>4.1313</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.472)</td>
<td>(0.4527)</td>
<td>(0.537)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5437</td>
<td>0.4897</td>
<td>0.558</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.2092)</td>
<td>(0.2068)</td>
<td>(0.2404)</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>288.22</td>
<td>288.05</td>
<td>514.66</td>
</tr>
</tbody>
</table>

Note: MSM estimation for entire group in levels and first differences. Cell size is 43031. The last row reports a test of the overidentifying restrictions distributed as a Chi-squared with 38 and 37 degrees of freedom respectively. The critical values at 5% are 53.10 and 51.91. Efficient estimates are reported in the second column with a weighting matrix $\Omega$ computed from the first step.

we to present figures which showed both our fitted values and the data unadjusted for family size and cohort effects, the CEQ LCH would look better and our procedure less naive.

Next we use our structural model to estimate both the discount rate and the coefficient of relative risk aversion. The resulting estimates are reported in Table 5.5.\textsuperscript{49}

Estimation in levels yields tight estimates of both parameters of interest.\textsuperscript{50} The structure which our model imposes on the data gives us strong predictions about preferences. Although first-difference estimates are slightly less precise, they do not appear to contradict the level estimates. Moreover, the overidentifying restrictions are more strongly rejected for the difference method. Therefore, in what follows, we will refer

\textsuperscript{49}MSM estimation is performed in 3 steps. First, we do a broad grid search over the parameter space, then a finer grid search and lastly use an optimization routine. This ensures that the program identifies the global minimum. The programs are CPU-intensive: one evaluation of the objective function requires approximately 45 minutes on an experimental P6 (or "686") chip running at 133 Mhz (courtesy of Intel Corp.). A $25\times25$ grid search therefore requires 12 days. For details, see Appendix A.

\textsuperscript{50}An earlier version of this chapter presented pictures of our objective function which demonstrated that the optima lay in a valley in $(\beta, \rho)$ space. Many interpreted this as implying that the separate identification of $\beta$ and $\rho$ was tenous— that we could really only accurately estimate a linear relation between parameters. We want to emphasize that the valley picture is correct, but that the valley has a clear minima which gives us estimates of both parameters.
only to level estimates. Efficient estimation has a minor impact on the point estimates or their standard deviations. In both level and first difference estimation, we can reject strongly the overidentifying restrictions. The 5% critical value for a $\chi^2$ (38) is 53.10, and 51.91 for a $\chi^2$ (37). This is not entirely surprising, given the number of moments we use. Moreover, the initial and last moments (average consumption at young ages and just before retirement) are each possibly misspecified if we have misspecified the initial distribution of cash on hand or the retirement rule.

Our estimates of the discount rate are close to the interest rate. In the efficient estimation case, we cannot reject the hypothesis that the two are equal at standard levels of confidence. It is worth noting that the discount factor which we estimate is within a reasonable range. Using information on the elasticity of assets with respect to uncertainty, Carroll and Samwick (1994) estimate that discount rate is in the vicinity of 10-15% and argue that even higher discount rates are needed to rationalize the findings of Hubbard et al. (1994). Our lower discount rate, however, does not imply that households are not impatient enough to generate buffer stock behavior. Lower levels of impatience generate buffer stock behavior when combined with steep income profiles.

With our estimates in hand, we can now address how well the stochastic model fits the life-cycle consumption profile. Figure 5.4 plots the simulated and actual consumption data along with the income profile, and a 95% pointwise confidence interval for the simulated profile. The stochastic life-cycle model does a much better job at fitting the consumption profile than the certainty line. Consumption tracks income until around age 40 – 42 and then falls sharply, as the household starts building up its retirement savings. Simulated consumption never exceeds income, except in the first periods of life. The tight structure imposed by the model is able to deliver good predictions in terms of consumption dynamics, despite having only two free parameters to work with.

---

51 Unfortunately, the two cases we analyzed are not nested, so that hypotheses testing is not possible.
52 This result is simply an artifact due to our assumptions regarding the initial level of cash on hand.
Why are we able, within the context of our model to obtain such tight estimates of the discount rate? Figure 5.5 plots various simulated profiles for different $\beta$ between 0.91 and 0.95, corresponding to a discount rate between 4.71 and 9.98 percent. It is immediately obvious that the profiles are quite sensitive to this parameter. With a higher discount factor, the agent is willing to save more and earlier for retirement purposes. The consumption path exhibits less of a hump shape, and may even be increasing over the entire working life. On the other hand, for more impatient consumers, consumption parallels income until much later in life and then falls more precipitously to build assets for retirement. This implies a stronger concavity of the consumption profile. Thus, our method will yield tight estimates of the discount factor precisely because the discount factor drives the hump shape in consumption.

We now turn to the question of how household behavior changes over the life cycle. Figure 5.6 displays the average household saving rate when there are no initial assets ($x_0 = 0$). Two distinct phases in the consumer’s life are visible. Until about age 40, households built their buffer and then consume roughly their income. Around age 40, retirement considerations induce an increase in saving. From then until the age of retirement, households consume much less than their current income.

We can put some additional structure on these two phases by looking at the target level of cash on hand at each age, as defined in section 2. Figure 2.3, already reported, plots next period expected cash on hand as a function of current cash on hand for various ages. Figure 5.7 directly computes the target level of cash on hand for consumers aged 25 – 48. One can see from the graph that the target level of cash on hand remains small early in life, around 1.3 times permanent income. Shortly after 40, the target increases substantially, as consumers try to build their retirement nest-egg. This figure shows a dramatic change in behavior. When the target level of liquid wealth is small, agents are “buffer stock”. Their consumption closely follows their income. Around age 41, agents
desire to accumulate assets for retirement. With a large stock of wealth relative to permanent income, consumers can smooth high frequency movements in income. Their behavior more closely mimics that of certainty-equivalent consumers.

None of these results are assumed in our model. If we had found either consumers to be more patient or flatter adjusted income profiles, households could have behaved as life-cycle consumers for their entire lives. Similarly, very impatient consumers could have been buffer stock consumers all of their working lives, relying on illiquid wealth to finance consumption during retirement.

We can also decompose total saving at each age into life-cycle and buffer-stock saving. Our previous discussion might lead the reader to think that agents have no concern for retirement when they are young and no concern for labor income uncertainty later in life. This is incorrect since consumers are rational and perfectly foresee their retirement needs. To proceed, we define total saving as the discounted variation in financial wealth from one period to the next, using our simulated profile:\textsuperscript{53}

\[ S_t = (W_{t+1} - W_t) / R = (R - 1) / R \cdot W_t + Y_t - C_t. \]

Saving is equal to investment plus labor income minus consumption, i.e. to disposable income minus consumption. Next, for the estimated parameters, we compute the consumption path \( \{C_t^{LC}\} \), that would occur under certainty.\textsuperscript{54} We then define life-cycle saving as the difference between total income and life-cycle consumption

\[ S_t^{LC} = (W_{t+1}^{LC} - W_t^{LC}) / R = (R - 1) / R \cdot W_t^{LC} + Y_t - C_t^{LC} \]

and buffer stock saving is defined as the complement. Figure 5.8 plots the precautionary

\textsuperscript{53}The discount comes from our assumption that income is received and consumption occurs at the beginning of the period. See (2.2).

\textsuperscript{54}In order to do this, we calibrate the certainty case, so as to yield the same consumption rule at retirement. Consumers effectively have strong bequest motives under these assumptions.
saving, liquid and total life-cycle saving. The latter is defined by adding back to income pension and social security contributions. Given the estimated discount rate, CEQ life-cycle consumers would like to borrow early in life. However, precautionary saving motives cause them to hold a positive buffer stock of wealth. Around age 40, in accordance with our previous characterization, life-cycle savings becomes larger than precautionary savings. The need to build retirement savings sets in. As asset levels increase, the expected variance of consumption declines, decreasing the precautionary saving motive. This latter effect, which our previous decomposition masked, induces the agent to run down the buffer. As a result, the total saving rate later in life is smaller than under the certainty equivalent framework.

5.2 Disaggregated evidence.

We next fit the model to each occupation and education cell separately. To begin with, we present the results for the CEQ LCH case across education and occupation groups in Table 5.6. Note that here and subsequently, we do not incorporate possible cross-cells correlation.\footnote{Note also that the same individuals are allocated both to an education and an occupation cell. Therefore, results across education/occupation are not independent.}

Again, the results indicate that the discount rate is accurately estimated and is roughly equal to the interest rate. Higher education levels tend to have a higher discount factor. Figures 5.2 and 5.3 present these fitted profiles. One can see the increase in the slope for higher educational groups.

To estimate across cells using our stochastic model, we simply follow the procedure described above. However, due to computing power constraints, we are unable to search over the entire $(\beta, \rho)$ space for all cells. Since the results are more sensitive to the discount factor, we fix the intertemporal elasticity of substitution at its aggregate value, $1/0.49 = 2.04$, and search across values of the discount rate. Each cell's optimization is
run with a different income profile (Figures 3.8 and 3.9), income uncertainty (Table 5.3),
and retirement consumption rules (Table 5.4), while we impose a constant probability
of zero income ($p = 0.5\%$) and real interest rate ($R = 1.03$). The results are summarized
in Table 5.7. Except for the last occupation cell (Self-Employed), the parameters are
remarkably close to those estimated using the aggregate profile. As we observed with
the benchmark case, the discount rate decreases weakly with education and ranges from
3.31\% to 4.25\%. The associated fitted profiles are displayed in Figures 5.9 and 5.10. The
fit is quite good for most cells, except Self-Employed and College Graduates.\footnote{Note that for all cells the standard errors are smaller than for the aggregate estimation. This simply reflects the fact that $p$ is fixed. The correlation between $p$ and $\beta$ in the aggregate estimation increases the standard errors. We hope in the future to be able to estimate both parameters for each cells.} For all
groups except the first educational group, the consumption profile is humped. As in the
aggregate case, we reject the overidentifying restrictions in all cases. We conclude, given
the similarity in the estimates that our results are quite robust to heterogenous shocks
and processes.
Table 5.7: Estimates from the Stochastic Model, $\rho = 0.4897$

<table>
<thead>
<tr>
<th>Group</th>
<th>$\beta$</th>
<th>S.E.</th>
<th>$\delta$</th>
<th>S.E.</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some Highschool</td>
<td>0.9607</td>
<td>(1.33 $10^{-5}$)</td>
<td>4.09</td>
<td>(0.001)</td>
<td>256.73</td>
</tr>
<tr>
<td>Highschool Graduate</td>
<td>0.9592</td>
<td>(2.84 $10^{-5}$)</td>
<td>4.25</td>
<td>(0.003)</td>
<td>107.41</td>
</tr>
<tr>
<td>Some College</td>
<td>0.9629</td>
<td>(2.37 $10^{-4}$)</td>
<td>3.84</td>
<td>(0.025)</td>
<td>85.25</td>
</tr>
<tr>
<td>College Graduate</td>
<td>0.9679</td>
<td>(1.62 $10^{-5}$)</td>
<td>3.31</td>
<td>(0.002)</td>
<td>115.36</td>
</tr>
<tr>
<td>Graduate School</td>
<td>0.9656</td>
<td>(2.00 $10^{-5}$)</td>
<td>3.56</td>
<td>(0.002)</td>
<td>165.96</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial and Prof.</td>
<td>0.9621</td>
<td>(1.75 $10^{-4}$)</td>
<td>3.94</td>
<td>(0.018)</td>
<td>118.11</td>
</tr>
<tr>
<td>Tech., Sales, Admin.</td>
<td>0.9669</td>
<td>(2.49 $10^{-5}$)</td>
<td>3.42</td>
<td>(0.003)</td>
<td>74.92</td>
</tr>
<tr>
<td>Precision Prod., Craft</td>
<td>0.9641</td>
<td>(4.60 $10^{-5}$)</td>
<td>3.72</td>
<td>(0.005)</td>
<td>123.43</td>
</tr>
<tr>
<td>Operators, Laborers</td>
<td>0.9655</td>
<td>(4.29 $10^{-5}$)</td>
<td>3.57</td>
<td>(0.005)</td>
<td>36.53</td>
</tr>
<tr>
<td>Self Employed</td>
<td>0.9555</td>
<td>(6.57 $10^{-3}$)</td>
<td>4.66</td>
<td>(0.719)</td>
<td>67.09</td>
</tr>
</tbody>
</table>

Note: MSM estimation in levels over $\beta$. Cell size given in Table 5.1. The last column reports a test of the overidentifying restrictions distributed as a Chi-squared with 38 degrees of freedom. The critical value at 5% is 53.10.

Further, as in the aggregate case, the estimated discount factors are consistently lower in our estimation of the stochastic model, than in the CEQ LCH baseline case. We can reject that the interest rate and the discount rate are equal in all cells, except Self-Employed.

5.3 Robustness Checks and Extensions.

Our estimation procedure depends on the calibrated parameters, $\chi$. In this section we investigate the robustness of our results to these parameters. Due to computing constraints, we have only been able to check the robustness with respect to the discount factor. In what follows, we again maintain a constant intertemporal elasticity of substitution equal to 2.04. The results are reported in Table 5.8.

The estimate of the discount rate is, as mentioned, most sensitive to our choice of interest rate. This is not surprising, since, early in life the difference between the interest rate and the discount rate is a key determinant of whether consumers exhibit buffer stock...
Table 5.8: Robustness Checks, \( \rho = 0.4897 \)

<table>
<thead>
<tr>
<th></th>
<th>( R = 1.05; \gamma_0 = 0.493; \gamma_1 = 0.035 )</th>
<th>( \beta )</th>
<th>S.E.</th>
<th>6</th>
<th>S.E.</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p = 0.05; )</td>
<td>0.9441</td>
<td>(3.28 ( 10^{-4} ))</td>
<td>5.92</td>
<td>(0.036)</td>
<td>266.25</td>
</tr>
<tr>
<td>2</td>
<td>( \sigma_u^2 = 0; \sigma_n^2 = 4.79 ( 10^{-4} ) )</td>
<td>0.9631</td>
<td>(8.56 ( 10^{-5} ))</td>
<td>3.83</td>
<td>(0.009)</td>
<td>275.07</td>
</tr>
<tr>
<td>3</td>
<td>( \gamma_0 = 0.535; \gamma_1 = 0.074 )</td>
<td>0.9612</td>
<td>(5.4 ( 10^{-5} ))</td>
<td>4.03</td>
<td>(0.006)</td>
<td>471.92</td>
</tr>
<tr>
<td>4</td>
<td>( x_0 = 0.0 )</td>
<td>0.9623</td>
<td>(1.9 ( 10^{-4} ))</td>
<td>3.91</td>
<td>(0.020)</td>
<td>489.96</td>
</tr>
<tr>
<td>5</td>
<td>Large Income; ( \gamma_0 = 0.241; \gamma_1 = 0.035 )</td>
<td>0.9899</td>
<td>(0.1676)</td>
<td>1.01</td>
<td>(17.10)</td>
<td>30596</td>
</tr>
</tbody>
</table>

Notes: MSM estimation in levels over \( \beta \). Cell size is 43031. The last column reports a test of the overidentifying restrictions distributed as a Chi-squared with 38 degrees of freedom. The critical value at 5% is 53.10. (1) assumes an interest rate of 5% and recomputes accordingly the last working period consumption rule; (4) computes the last working period consumption rule using asset data from Venti and Wise (1993). Large income in (6) includes mortgage payments and pension contributions.

behavior. Late in life, consumers will behave in a manner more consistent with the CEQ LCH, in which the change in consumption is driven by the product \( \beta R \).

So far we assumed a real interest rate of 3% a year. Although this is close to the long run real rate on liquid assets, some savings are held in longer term assets, yielding on average higher returns. Therefore, we re-estimate the discount rate assuming a real interest rate of 5\%.\(^{57}\) The estimate of the discount factor is now \( \beta = 0.9441 \), with a standard deviation of \( 3.28 \times 10^{-4} \). This implies a discount factor of 5.92%, almost exactly two hundred basis points higher than our previous estimate. Thus it is important to note that our results do not provide a tight estimate of the discount factor per se. However, our model gives precise estimates of the difference between the interest rate and the discount rate. This indicates that we capture mostly the substitution effect. The simulated profile reported in Figure 5.11.1 is roughly similar to our benchmark case, although consumption is smaller later in life to reflect the smaller \( \gamma \).

We next check the robustness of our results to the probability of zero income, \( p \). A

\(^{57}\)With a 5% interest rate, we recompute the last period consumption rule as \( \gamma_0 = 0.493; \gamma_1 = 0.035 \). Thus illiquid assets are more important and consumption is less sensitive to current cash on hand.
higher $p$, by increasing uncertainty, should lead to a larger buffer stock, implying less lifecycle savings late in life. Thus, an increase in $p$ is likely to yield a higher discount rate, to counteract the increase in precautionary savings. Our results indicate that this effect is quite weak. When $p = 5\%$, a tenfold increase, the discount rate actually decrease marginally to 3.83\%. On the other hand, given unemployment benefits, government assistance programs, one might argue that households can never experience zero income. Decreasing $p$ has even smaller effects on the estimated discount rate. Looking at the simulated profile, we see that the change in $p$ affects mostly consumption early in life. After age 40, asset level can buffer transitory fluctuations in income.

Next, we investigate the sensitivity of our results to the variance of permanent and transitory components, $\sigma^2_u$, and $\sigma^2_p$. We decreased the uncertainty faced by the agent from both sources of shocks. Intuitively, this should lead to a smaller buffer and should give a lower estimate of the discount rate. Our test eliminates altogether transitory shocks and reduces the variance of the permanent shocks to $\sigma^2_p = 4.8 \times 10^{-3}$. Our estimated discount rate, 3.81\%, confirms our intuition and reemphasizes that our results are reliant upon the underlying individual uncertainty. Looking at the simulated profile, we see that consumption is higher early in life and lower later, as the lower buffer translates into smaller accumulated assets.

An important assumption of our model concerns the retirement rule. As described in the previous section, the consumption rule at retirement is calibrated from CEX data, using both income and asset reports. To the extent that asset data are not accurate, our consumption rule is likely to be mismeasured. This, in turn, will affect the lifecycle profile as we near retirement. In order to test robustness to our hypothesis, we calibrate the retirement consumption rule using asset data reported in Venti and Wise (1993). The resulting values are $\gamma_0 = 0.535$; $\gamma_1 = 0.074$. Looking at (4.21), this indicates both a substantially smaller asset accumulation (as $\gamma_1$ increases) and a larger share of illiquid assets (as $\gamma_0$ increases). With a larger $\gamma_0$, the agent can rely on illiquid saving at
retirement time. This should lead to a smaller liquid asset accumulation and to a lower discount rate. A larger \( \gamma_1 \) has ambiguous effects. It increases the level of consumption out of cash on hand in the last working period. This is compatible both with a higher discount rate (as the household will consume more of a given wealth) and a lower discount rate (if the households accumulates more wealth). Thus, the actual consumption profile may indicate more or less preference for the present. We estimate a discount rate of 4.03\%, slightly higher than the benchmark estimate. This suggests that the presence of illiquid assets play an important role in the precision of our estimates and that the increase in \( \gamma_1 \) dominates. Looking at Figure 5.11.4, we see that assumptions about the last working period consumption rule affect the simulated profile substantially.

We then check the validity of the assumption \( x_0 = 0.3 \). As for the retirement rule, this is likely to affect our estimates by shifting the consumption profile at young ages. We reestimate our aggregate problem assuming that \( x_0 = 0 \). The estimated \( \beta \), 0.9623, is extremely close to our original one. With a lower initial level of cash on hand, households build theirs buffers early in life. This results in slightly lower consumption for the first few years. However, from then on, the consumption profile is similar. Thus, this assumption does not have a large impact on our estimates.

Our estimation procedure is also extremely sensitive to our assumed income profile. Permanent income growth is a key variable and determines to a large extent buffer-stock behavior, as we have demonstrated. Our definition of income subtracts pension contributions and mortgage payments, as they are likely to reflect illiquid saving, for which our theory is ill-equipped. However, it is possible in some circumstances to draw on voluntary pension contribution. Similarly, housing wealth is not entirely illiquid. Thus we reestimate our model assuming that these components of income contribute to liquid savings instead of illiquid savings. The associated values of the consumption rule at retirement are \( \gamma_0 = 0.241; \gamma_1 = 0.035 \). Since pension contribution are now part of liquid savings, the accumulated illiquid savings and \( \gamma_0 \) are lower. In effect, this amounts
to shifting upwards the income profile and increases substantially measured savings. Not surprisingly, this can only be matched by a lower discount rate. The estimated discount rate is 1.01%, well below the interest rate. The parameters are poorly estimated and the overidentifying restriction is enormous. This increase in savings can only be matched by assuming that consumers are extremely patient. However, this fails to capture the life-cycle pattern of consumption: the simulated profile grows exponentially.

We conclude that, except with respect to our definition of liquid income, our estimates and inference are reasonably robust to the calibration of our model.

6 Conclusion.

Macroeconomic models generally represent the consumer as an infinitely-lived, rational, representative agent, who behaves in accordance with the Permanent Income Hypothesis. Some analyses provide explicit microeconomic justification for this assumption, by deriving an insurance system which protects individuals from any idiosyncratic consumption risk (e.g. Rogerson (1988)) or by assuming that individual budget constraints never bind, so that aggregate behavior mimics individual behavior (e.g. Barro (1974)). Other models incorporate certain forms of individual heterogeneity when aggregating (e.g. Blanchard (1985)). The resulting representative agent will not, in general, have the same characteristics as individual consumers.

However, nearly all currently employed macroeconomic models include some form of a representative agent facing representative shocks. Such a representative agent framework constitutes an extremely powerful tool, mainly because the restrictions that optimizing behavior place on individuals will carry over to the aggregate economy. In this chapter, while we work in an explicitly partial equilibrium environment, we find important and substantial deviations from the canonical representative consumer model.

Using individual-level data to construct average profiles of income and consumption
over the working lives of households, we demonstrated that consumption remains hump-shaped, even after controlling for family and cohort effects. We then developed a model of consumption behavior embedding realistic levels of income uncertainty and estimated individual consumption functions using the Method of Simulated Moments. The model fits well and yields tight estimates of the discount rate and intertemporal elasticity of substitution. To the best of our knowledge, this is the first chapter which uses explicit individual uncertainty and the life-cycle profile of consumption to identify structural parameters of the utility function. The results indicate that consumers hold only a buffer stock of liquid assets in order to offset labor income fluctuations, until around age 42. Then, they start saving actively for retirement purposes. These two phases in consumer behavior are quite distinct and are at the heart of our identification procedure.

This age-heterogeneity of consumer behavior has important implications for aggregate consumption. In particular, it may rationalize Campbell and Mankiw (1989) finding that roughly 40% of all agents are "hand to mouth". In our interpretation, this may simply reflect the consumption of young households. In later work, we plan to investigate in more detail the aggregate implications of the age-heterogeneity of consumers.
Appendix D

a  Solving the consumer's problem.

In this appendix we describe our approach to solving numerically the consumer problem.

a.1  Euler equation.

The algorithm exploits the recursive structure of the consumer problem by solving the Euler equation. Given a consumption rule at age $t+1$, $c_{t+1}(\cdot)$, the algorithm solves for the consumption rule $c_t(x_t)$ that satisfies for any $x_t$:

$$u'(c_t(x_t)) = \beta R \mathbb{E}[u'(c_{t+1}(x_{t+1}))]$$

$$= \beta R \left( p \mathbb{E}[u'(c_{t+1}(x_{t+1})) | U_{t+1} = 0] + (1 - p) \mathbb{E}[u'(c_{t+1}(x_{t+1})) | U_{t+1} > 0] \right).$$  \hspace{1cm} (a.1)

a.2  Gauss-Hermite quadrature.

Assume for the time being that we know how to compute $c_{t+1}(\cdot)$ for all values of cash on hand. Our first problem consist in evaluating the expectation in (a.1). One can rewrite the Euler equation using the Intertemporal budget constraint,

$$x_{t+1} = X_{t+1}/P_{t+1} = (x_t - c_t) \frac{R}{G_{t+1}N_{t+1}} + U_{t+1},$$  \hspace{1cm} (a.2)

as:

251
\begin{align*}
    u'(c_t(x)) &= \beta R \left( p \mathbb{E}\left[ u' \left( c_{t+1} \left( (x - c_t) \frac{R}{G_{t+1} N} \right) G_{t+1} N \right) \right] + \\
    &\quad (1 - p) \mathbb{E}\left[ u' \left( c_{t+1} \left( (x - c_t) \frac{R}{G_{t+1} N} + U \right) G_{t+1} N \right) \mid U > 0 \right] \right).
\end{align*}

Since \( N \) and \( U \) are log normally distributed, the natural way to evaluate these integrals is to perform a two dimensional Gauss-Hermite quadrature:

\begin{align*}
    E[u'(c_{t+1}(x_{t+1})G_{t+1}N)] &= \int u' \left( c_{t+1} \left( (x - c_t) \frac{R}{G_{t+1} N} + U \right) G_{t+1} N \right) dF(N) \; dF(U) \\
    &= \int_{-\infty}^{\infty} f_1(n,u) e^{-n^2} e^{-u^2} \; du \; dn \\
    &= \sum_{i,j} f_1(n_i,u_j) \omega_{ij},
\end{align*}

where \( f_1(n,u) = \frac{1}{\pi} u' \left( c_{t+1} \left( (x - c_t) \frac{R}{G_{t+1}} e^{-\sqrt{2}\sigma_n n} + e^{-\sqrt{2}\sigma_u u} \right) G_{t+1} e^{-\sqrt{2}\sigma_n n} \right) \). The weights \( \omega_{ij} \) and nodes \( n_i, u_j \) are tabulated in Judd (1993). In practice, we performed a quadrature of order 12.

One can then find the root of the Euler equation at any point \( x \) using a standard Newton method. In practice, we constrain the root to be positive and less than \( x \), the current cash on hand. As discussed in the text, this restriction is always satisfied when there are no illiquid assets. Since illiquid assets cannot be borrowed against, it is also satisfied in their presence.

### a.3 Consumption rules.

We initialize the algorithm with the consumption rule at retirement: \( c_T(x_T) = \gamma_0 + \gamma_1 x_T \). One can show that the consumption rules for this problem are continuously differentiable as long as there are no liquidity constraints. However, in the presence of liquidity constraints, the consumption rules may exhibit a kink. See Deaton (1991) and Ayagari (1993). We effectively impose a liquidity constraint by not allowing the household to borrow against illiquid assets. This indicates that smooth approximation methods, as advocated by Judd (1992) are inappropriate in that case. Instead, we will use a standard discretization method: we specify an exogenous grid for cash-on-hand: \( \{x^j\}_{j=1}^J \subset [0, x^{\text{max}}] \). In order to capture the curvature of the consumption
rule at low values of cash on hand, the grid will be finer for $x \in [0, 2]$. In practice, for each value of cash on hand on the grid, $x'$, we find the associated consumption $c'$ that satisfies (a.1). In choosing the size and coarseness of the grid, we face the usual trade-off between precision and computing time. Adding points on the grid gives a finer approximation of the consumption rules. Since the consumption rule at age $t + 1$ is the input necessary to get the consumption rule at age $t$, imprecisions could compound over time. On the other hand, the Euler equation is the innermost loop of the entire algorithm. With 100 points on the grid and 40 time period, we must solve 4000 solutions to (a.1). This takes approximately 45 minutes on a P6 chip running at 133 MHz, courtesy of Intel Corporation. We also face a difficult decision regarding the range of cash on hand, $x^{max}$. For small values, cash on hand in sample is likely to move out of the grid. Consumption will then be evaluated using extrapolation methods, always much less precise than interpolation. On the other hand, increasing the range with a fixed number of points implies less precise estimation of the curvature. One solution consists in endogenizing the grid so that, for instance, cash on hand remains within the grid with probability 0.95. We adopted the simpler approach consisting in checking that cash on hand, in the simulations, remains strictly inside the grid. In practice, we took $x^{max} = 40$ and $J = 100$, with 50 points between 0 and 2. We checked the quality of the approximation by solving the stationary infinite horizon problem and checking the rate of convergence to the fixed point of the functional Bellman equation.

b Data.

We use the CEX family, member and detailed expenditure files for years 1980 to 1993, as kindly provided by the NBER. Most of our information about the CEX is obtained from Bureau of Labor Statistics (1993, and years 1980-1992) and conversations with DLS statisticians. Households are discarded if they are missing any of the information necessary for the regressions, if they report changes in age from the second to fifth interview of more than a year or less than zero years, if they are classified as incomplete income reporters, or if their reporting implies less than 1000 in annual income or consumption.

We use information about the reference person to assign the household to cells, unless the reference person is female. In this case we use the spouses information. If there is no spouse, or his information is missing, the household is discarded. When this cut was made it eliminated
20% of the sample. All information besides individual labor income and consumption is taken from the family files. Values are assigned to a household based on information gathered in the fifth interview, otherwise information is used from the second interview, or, if it is not available, the household is discarded. Households should not be matched across 1985 to 1986, and are not. Care is taken to assure consistency in our data despite variable classification changes through time, and across reference person and spouse. Information was kindly provided by the Division of the CEX in the Bureau of Labor Statistics about various issues including the matching of occupation codes from 1980-81 to later years.

Pension contributions, income, Social Security contributions, and all asset income all refer to the past twelve months. Our definition of pension contributions is the sum over the CEX subcategories and thus includes private pensions, public pensions, Railroad Retirement pensions, and self-employed, IRA, and Keogh plans. If the after-tax family income variables is topcoded, reference person and spouse labor incomes are subtracted and we add, for each, the variable created by multiplying the earnings in last paycheck by the appropriate pay period. These labor income variables are the sole variables from the member files used. Assets and asset income refers to the sum over savings accounts, checking accounts, bonds, and stocks, as of the time of interview. Each household is assigned to a year based on the midpoint between the first and fifth interview if both data are available; otherwise simply the single interview date is used. Age is the average of both interviews if both are available, otherwise it is the single one available. Due to some extreme reports, we reset reported tax rates above 50% back to 50%, and below zero percent to zero. We perform a similar exercise for Social Security contribution rates and pension contribution rates, using 25% as the upper bound.

Consumption data is compiled from the detailed expenditure files as all expenditures by a household except for those for health care, mortgage interest, and education. The consumption level is then the average monthly expenditure times twelve. Five percent of households have consumption data for 4, 7, 10, 13, or 14 months and these households' consumption are treated as if they were over 3, 6, 9, and 12 months. That is the recall interview period extended beyond the basic three months and some expenditures are recorded in a later month. BLS statisticians recommend treating these expenditures as if they occurred in the preceding month. Those covering 1 or 2 months (one percent of the sample) were dropped.

The unemployment rates merged to the CEX are the regional unemployment rates for civilian

c Method of Simulated Moments (MSM).

According to section 2, consumption at age \( t \) for individual \( i \) depends on cash on hand \( x^i_t \), the realization of permanent component of income \( P^i_t \), the entire path of expected permanent income growth \( G_t = \{G_t\}_{t=1}^T \), and the parameters of the consumption problem \( \hat{\theta}' = (\beta, \rho, R, \gamma_0, \gamma_1, p, \sigma_u^2, \sigma^2) \), an 8x1 vector. In practice, it will not be possible to estimate directly all the elements of \( \hat{\theta} \). Instead, we rewrite \( \hat{\theta} = (\theta', \chi)' \) where \( \theta = (\beta, \rho)' \) and \( \chi = (R, \gamma_0, \gamma_1, p, \sigma_u^2, \sigma^2)' \).

We assume that \( \theta \) belongs to some compact set \( \Theta \subset \mathbb{R}^2 \). The elements of \( \chi \) will be inputs into the estimation procedure. Defining the vector of state variable \( z^i_t = (x^i_t, P^i_t) \), we postulate the following data-generating process:

\[
\ln C^i_t = \ln C_t (x^i_t, \theta; \chi, \Theta_T) + \epsilon^i_t = \ln(c_t (x^i_t, \theta; \chi, \Theta_T) P^i_t) + \epsilon^i_t, \tag{c.3}
\]

where \( \epsilon^i_t \) is an idiosyncratic shock that represents measurement error in consumption levels. We are interested in estimating \( \theta \). If we were able to observe simultaneously the level of cash on hand of consumers and the level of their permanent component of income (as well as their consumption), \( \theta \) could be estimated using Hansen's GMM (1982) on individual level data. More precisely, one would write the following moment conditions:

\[
E [h (w', \theta_0; \chi, \Theta_T)] = 0, \tag{c.4}
\]

where \( \theta_0 \) is the true parameter vector, and

\[
w^i_t = (\ln C^i_t, z^{ii}_t)',
\]

\[
h_t (w^i_t, \theta; \chi, \Theta_T) = (\ln C^i_t - \ln C_t (z^i_t, \theta; \chi, \Theta_T)) \frac{\partial \ln C_t (z^i_t, \theta; \chi, \Theta_T)}{\partial \theta}.
\]

\( w \) is a \( T \times 3 \) vector and \( h (w, \theta, \Theta_T) \) is a \( T \times 2 \) vector.

The estimation procedure would then minimize:
\[ g(\theta; \chi, \Theta)^T W g(\theta; \chi, \Theta), \]  
\hspace{1cm} (c.5) \]

where \( g(\theta; \chi, \Theta) = \text{vec} \left( \frac{1}{I} \sum_{i=1}^{I} h \left( w^i; \theta; \chi, \Theta \right) \right) \) is a \( 2T \times 1 \) vector and \( W \) is a weighting matrix. In practice, the number of cross-section observations for each age varies in the sample, so that \( I = I_t \). We do not write explicitly this extra time-dependence in order to keep the notations simpler. The first difficulty with (c.4) is that quality panel data on consumption, asset and income information for individual households are not available in any US dataset. Therefore direct estimation using (c.5) is not possible. We do however observe \( \chi \) and average consumption at each age \( \ln \hat{C}_t \equiv \frac{1}{T} \sum_{i=1}^{T} \ln C^i_t \). This suggests that we can circumvent the problem by looking directly at the unconditional expectation of consumption at each age:

\[ \ln \hat{C}_t (\theta; \chi, \Theta) \equiv E[\ln C_t (z_t; \theta; \chi, \Theta)] = \int \ln C_t (z, \theta; \chi, \Theta) \ dF_t (z), \]  
\hspace{1cm} (c.6) \]

where the unconditional distributions of normalized cash on hand and permanent income depend on age \( t \). We write the \( T \) moment conditions as

\[ E \left[ h \left( \ln C^i, \theta_0; \chi, \Theta \right) \right] = 0, \]  
\hspace{1cm} (c.7) \]

where \( \theta_0 \) is the true parameter vector, \( \ln C^i = \{ \ln C^i_t \}_{t=1}^{T} \) and \( h \left( \ln C^i, \theta; \Theta \right) \) is a \( T \times 1 \) vector with \( t^{th} \) element:

\[ h_t \left( \ln C^i, \theta; \chi, \Theta \right) = \ln C^i_t - \ln \hat{C}_t (\theta; \chi, \Theta). \]

However, at this point we encounter a second difficulty. The unconditional distribution for the state variables at age \( t \), \( dF_t (z) \), is extremely cumbersome to evaluate, as well as the unconditional expectation (c.6).

The Method of Simulated Moments, as developed by Pakes and Pollard (1989) and Duffie and Singleton (1993) allows us to circumvent this difficulty. We can define a measurable transition function \( \Phi: \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^T \times \Theta \rightarrow \mathbb{R}^2 \) that describes the dynamics of the state variables \( z_{t+1} = (x_{t+1}, \nu_{t+1}) = \Phi (z_t, \nu_{t+1}, \Theta, \theta) \), where \( \nu_t = (U_t, N_t) \) according to:

\[ x_{t+1} = \frac{R}{G_{t+1} N_{t+1}} (x_t - c_t (x^i_t, \theta; \Theta)) + U_{t+1} \]  
\hspace{1cm} (c.8) \]
\[ P_{t+1} = G_t P_t N_{t+1}. \]

We omit the dependence in the calibrated parameters \( \chi \). \( U_{t+1} \) and \( N_{t+1} \) are respectively the permanent and transitory shocks to income. The first line of (c.8) is the normalized budget equation while the second line follows from our assumptions about the income process. This transition function can then be used to rewrite the unconditional expectation (c.6):

\[
\ln \hat{C}_t(\theta; \chi, \Theta) = \int \ln C_t(z, \theta; \chi, \Theta) \, dF(z) \quad \text{(c.9)}
\]

\[
= \int \int \ln C_t(\mathbf{T}(z, \nu, \Theta_T, \theta), \theta; \chi, \Theta) \, dF_{t-1}(z) \, dF(\nu).
\]

Note that the transition function depends on \( \theta \), through the consumption rule. (c.9) provides a convenient way to calculate the unconditional expectation is to use a Monte-Carlo integration. Assume that we have an \( \mathbb{R}^2 \times \mathbb{R}^T \)-valued sequence of random variables \( \{ \tilde{v}^i \}_{i=1}^{L} \) where \( \tilde{v}^i = (\tilde{v}_1, ..., \tilde{v}_T)' \), identically distributed and independent of \( \{ \nu^i \}_{i=1}^{L} \). From any initial distribution \( F(z_0) \) and candidate \( \theta \), we can generate the path of state variables according to (c.8):

\[
\tilde{z}_{t+1}^i = \mathbf{T}(\tilde{z}_t^i, \tilde{v}_{t+1}, \Theta_T, \theta); \quad \forall 1 \leq t \leq T \text{ and } 1 \leq i \leq L.
\]

For large enough \( L \), the unconditional expectation is then simulated by:

\[
\ln \hat{C}_t(\theta; \chi, \Theta_T) \equiv \frac{1}{L} \sum_{i=1}^{L} \ln C_t(\tilde{z}_t^i, \theta; \chi, \Theta_T) \sim \ln \hat{C}_t(\theta; \chi, \Theta_T).
\]

For any parameter vector \( \theta \in \Theta \) we can construct the \( T \) moments:

\[
\tilde{g}_t(\theta; \chi, \Theta_T) = \frac{1}{I(t)} \sum_{i=1}^{I(t)} \tilde{h}_t(\ln C_t^i, \theta; \chi, \Theta_T)
\]

\[
= \frac{1}{I(t)} \sum_{i=1}^{I(t)} \ln C_t^i - \ln \hat{C}_t(\theta; \chi, \Theta_T)
\]

\[
= \ln \hat{C}_t - \ln \hat{C}_t(\theta; \chi, \Theta_T).
\]

257
The estimation procedure minimizes:

\[
\tilde{g}(\theta; \chi, \mathfrak{G}_T)' W \tilde{g}(\theta; \chi, \mathfrak{G}_T),
\]

where \( \tilde{g}(\theta; \chi, \mathfrak{G}_T) = (\tilde{g}_1, ..., \tilde{g}_T)' \) is a T\times1 vector and \( W \) is a weighting matrix. Note that in the case where \( W = I_T \), the identity matrix, the estimation procedure is equivalent to minimizing the sum of square residuals:

\[
S(\theta; \chi, \mathfrak{G}_T) = \sum_{t=1}^{T} \left( \ln \hat{C}_t - \ln \hat{C}_t(\theta; \chi, \mathfrak{G}_T) \right)^2.
\]

However, even though we are minimizing the sum of squared residuals, asymptotic results still apply as long as \( I(t) \to \infty \) where \( I(t) \) is the number of observations at age \( t \). Under the regularity conditions stated in Pakes and Pollard (1989) and Duffie and Singleton (1993), the MSM estimator \( \hat{\theta} \) is both consistent and asymptotically normally distributed. Denoting \( I = \frac{1}{T} \sum_{t=1}^{T} I(t) \),

\[
\sqrt{I} \left( \hat{\theta} - \theta_0 \right) \to \mathcal{N}(0, V),
\]

with

\[
V = (D'WD)^{-1} D'W\Omega WD (D'WD)^{-1}
\]

\[
D = E[\partial \tilde{g}/\partial \theta']
\]

\[
\Omega = \text{avar}(\tilde{g})
\]

\[
W = \text{plim} W_I.
\]

In practice, the asymptotic variance-covariance matrix is estimated by:

\[
\hat{V} = \frac{1}{I} \left( \hat{D}'\hat{W}\hat{D} \right)^{-1} \hat{D}'\hat{W}\hat{D} \left( \hat{D}'\hat{W}\hat{D} \right)^{-1}
\]

\[
\hat{D} = \left. \frac{\partial \tilde{g}}{\partial \theta'} \right|_{\theta = \hat{\theta}}.
\]

258
The $t^{th}$ diagonal element of $\hat{\Omega}$ is given by

$$\hat{\Omega}_t = \frac{1}{f(t)} \sum_{i=1}^{f(t)} \left( \ln C_i - \ln \hat{C}_i (\theta; \chi, \Theta_T) \right)^2.$$ 

Under the assumption of no serial correlation, the off-diagonal elements of $\hat{\Omega}$ are 0. However, a robust estimator can be constructed if $\hat{\Omega}_{t,t'}$ is defined as:

$$\hat{\Omega}_{t,t'} = \left( \ln \hat{C}_t - \ln \hat{C}_i (\theta; \chi, \Theta_T) \right) \left( \ln \hat{C}_{t'} - \ln \hat{C}_i (\theta; \chi, \Theta_T) \right).$$

This methodology also provides a useful overidentifying restriction test. If the model is correct, the statistic

$$X_{T-2} = I \hat{g} (\theta; \chi, \Theta_T)' \hat{\Omega}^{-1} \hat{g} (\theta; \chi, \Theta_T)$$

is distributed asymptotically as Chi-squared with $T - 2$ degrees of freedom.

The optimal weighting matrix is $W = \Omega^{-1}$. The optimal weighting is implemented by first running the regression with an arbitrary weighting matrix, computing the associated $\Omega^{-1}$, and then using this estimate in a second round of estimation.

In practice, we simulate $\ln \hat{C}_i (\theta; \chi, \Theta_T)$ by running $L = 20,000$ independents income processes for 40 years, and computing in each year the associated consumption and cash on hand. The first step used $W = I_T$. We also assume that the initial distribution of cash on hand is 0.3 times current income. This assumption captures the fact that most households do not start with no assets. Finally, for all households, we set the initial value of the permanent component of income to the estimated income level at age 25 from our profiles. We performed first a 25x25 grid search over the parameter space $\Theta$. Then, we performed a second 25x25 grid search around the optimum. This guaranties that the procedure converges to the global minimum. Then, we used a standard minimization algorithm. Each grid search takes approximately 12 days of CPU time on a P6 or on a RSC6000. Once the optimum has been found, the gradient of the moment vector is evaluated numerically and the variance-covariance matrix estimated. For the disaggregated
results and the robustness checks, a Brent algorithm was used.
Figure 5.2.1: CONSUMPTION RULES

Figure 2.1.a: Consumption Rules

\[ \beta = 0.963, \rho = 0.490, \gamma_f = 1, G = 1 \]

Figure 2.1.b: Consumption Rules

\[ \beta = 0.963, \rho = 0.490, \gamma_f = 0.384, \gamma_t = 0.048, G = 1 \]
Figure 5.2.2: INDIVIDUAL PROFILES

Figure 2.2.a: Individual Profile

Figure 2.2.b: Individual Profile, 0-Income Shock
Figure 2.3: Expected Cash on Hand

\[ \beta = 0.963, \rho = 0.490 \]
Figure 5.3.1: HOUSEHOLD CONSUMPTION AND INCOME OVER THE LIFECYCLE
Figure 5.3.2: FAMILY SIZE
Figure 5.3.3: COHORT AND FAMILY ADJUSTMENT

Figure 3.3a
Consumption and Income With and Without Cohort Adjustment

Figure 3.3b
Consumption and Income With and Without Family Adjustments
Figure 5.3.4: TOTAL LIQUID ASSETS TO INCOME RATIO

Assets in Stocks, Bonds, Savings, and Checking Accounts

25  35  45  55  65

0  1  2  3  4
Figure 5.3.5: Consumption(0) and Income(+) Profiles by Education

A. Some High School
B. H.S. Degree
C. Some College
D. College Degree
E. Some Grad. Sch.
Figure 5.3.6: Consumption[0] and Income[+] Profiles by Occupation

Figure 3.6
Consumption[0] and Income[+] over the Lifecycle

A. Managerial, Professional

B. Tech, Sales, Admin Support

C. Precs. Prod, Craft, Repair

D. Operator, Fabricator, Lab.

E. Self-Employed
Figure 5.3.7: Rescaled Smoothed Consumption over the Lifecycle

Note: All series scaled to mean of unsmoothed total consumption
Total[-], Nondurable[0], and Food[+] Consumption
Figure 5.5.1: Income, Consumption and Consumption Predicted by CEQ LCH
Figure 5.5.2: By Ed: Income, Consumption, and Consumption Predicted by CEQ LCH

A. Some HighSchool

B. H.S. Degree

C. Some College

D. College Degree

E. Some Grad. Sch.
Figure 5.5.3: By Occ: Income, Consumption, and Consumption Predicted by CEQ LCH
Figure 5.5.4: SIMULATED AND ACTUAL CONSUMPTION PROFILES

Figure 5.4: Simulated and Actual Consumption Profiles
(confidence bands), $\beta=0.963$, $\rho=0.490$
Figure 5.5.5: Simulated Consumption Profiles for different $\beta$

$R=3\%$, $\rho=0.490$

The graph illustrates simulated consumption profiles for different values of $\beta$. The data points represent observed consumption, with lines indicating the simulated profiles for various $\beta$ values.
Figure 5.7: Target Cash on Hand

\[ \beta = 0.932, \rho = 1.573, R = 3\% \]
Figure 5.8: Life Cycle and Buffer Savings

$\beta = 0.963$, $\rho = 0.490$; $x_0 = 0$
Figure 5.5.11: Simulated and Actual Consumption Profiles, Robustness Checks

Figure 5.5.1: Simulated and Actual Consumption Profiles
\( \beta = 0.544, \rho = 0.490; R = 1.00 \)

Figure 5.5.2: Simulated and Actual Consumption Profiles
\( \beta = 0.563, \rho = 0.490; p = 0.52 \)

Figure 5.5.3: Simulated and Actual Consumption Profiles
\( \beta = 0.563, \rho = 0.490; \text{variance} \)

Figure 5.5.4: Simulated and Actual Consumption Profiles
\( \beta = 0.561, \rho = 0.490; \text{ Wise} \)

Figure 5.5.5: Simulated and Actual Consumption Profiles
\( \beta = 0.592, \rho = 0.490; \phi = 0 \)
Bibliography


Baily, Walter and Leo Kropywiansky, “A Test resolving the ‘Pr. 0 Problem’ in


Caballero, Ricardo, “Consumption Puzzles and Precautionary Saving,” Journal of


286


from Social Security Reform in Taiwan,” 1993. mimeo, Milken Institute for Job and Capital Formation.


Mussa, Michael, “Adaptive and Regressive Expectations in a Rational Model of the


**Serrat, Angel**, “Exchange Rate Dynamics in a Multilateral Target Zone,” December 1994. mimeo MIT.
Shea, John, “Union Contracts and the Life Cycle-Permanent Income Hypothesis,”

Souleles, Nicholas, “The Response of Household Consumption to Income Tax

Stokey, Nancy L., Robert E. Lucas, and Edward Prescott, Recursive Methods in

Svensson, Lars, “Target Zones and Interest Rates Variability,” Journal of

Venti, Steven and David Wise, “The Wealth of Cohorts: Retirement Saving and


Wessel, David, “Fed Was Split Over Rate Cut at July Meeting,” Wall street journal,
1995, August 28.


, “Rate Reversal: After Long Tightening, the Fed Loosens Up with a 1/4 Point

and Anita Raghavan, “Open Market: a Growing Glasnost at the Fed Is


Wyplosz, Charles, “Monetary Union and Fiscal Policy Discipline,” European

Zeldes, Stephen P., “Optimal Consumption with Stochastic Income: Deviations from