Essays on Finance, Learning, and Macroeconomics

by

Joseph Buchman Doyle, Jr.

Sc.B., Brown University (2007)

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

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Author ....
Department of Economics
August 15, 2012

Certified by ....
Anna Mikusheva
Castle-Krob Career Development Associate Professor of Economics
Thesis Supervisor

Certified by ....
Ricardo Caballero
Ford International Professor of Economics
Thesis Supervisor

Accepted by ....
Michael Greenstone
3M Professor of Environmental Economics
Chairman, Department Committee on Graduate Theses

Essays on Finance, Learning, and Macroeconomics

by

Joseph Buchman Doyle, Jr.

Submitted to the Department of Economics
on August 15, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of four essays on finance, learning, and macroeconomics.

The first essay studies whether learning can explain why the standard consumption-based asset pricing model produces large pricing errors for U.S. equity returns. I prove that under learning standard moment conditions need not hold in finite samples, leading to pricing errors. Simulations show that learning can generate quantitatively realistic pricing errors and a substantial equity risk premium. I find that a model with learning is not rejected in the data, producing pricing errors that are statistically indistinguishable from zero.

The second essay (co-authored with Anna Mikusheva) studies the properties of the common impulse response function matching estimator (IRFME) in settings with many parameters. We prove that the common IRFME is consistent and asymptotically normal only when the horizon of IRFs being matched grows slowly enough. We use simulations to evaluate the performance of the common IRFME in a practical example, and we compare it with an infrequently used bias corrected approach, based on indirect inferences. Our findings suggest that the common IRFME performs poorly in situations where the sample size is not much larger than the horizon of IRFs being matched, and in those situations, the bias corrected approach with bootstrapped standard errors performs better.

The third essay (co-authored with Ricardo Caballero) documents that, in contrast with their widely perceived excess return, popular carry trade strategies yield low systemic-risk-adjusted returns. In contrast, hedging the carry with exchange rate options produces large returns that are not a compensation for systemic risk. We show that this result stems from the fact that the corresponding portfolio of exchange rate options provides a cheap form of systemic insurance.

The fourth essay shows that the documented overbidding in pay-as-you-go auctions relative to a static model can be explained by the presence of a small subset of aggressive bidders. I argue that aggressive bidding can be rational if users are able to form reputations that deter future competition, and I present empirical evidence that this is the case. In auctions without any aggressive bidders, there is no evidence of overbidding in PAYGA.

Thesis Supervisor: Anna Mikusheva
Title: Castle-Krob Career Development Associate Professor of Economics

Thesis Supervisor: Ricardo Caballero
Title: Ford International Professor of Economics

Acknowledgments

I would like to begin by thanking my advisors, Anna Mikusheva and Ricardo Caballero, for generously dedicating their time and energy to my development as a researcher. Without their encouragement and support throughout my time at MIT, my experience would have been immensely more difficult, and I would not have come half as far as I have. I will continue to draw on their advice as I proceed further on my own.

I would also like to thank the others who have provided guidance on the research contained in this thesis. Those are Ivan Werning, Adrien Verdelhan, Hui Chen, Leonid Kogan, Gustavo Manso, Refet Gurkaynak, Barbara Rossi, Glenn Ellison, Panle Jia Barwick, Gregory Lewis, Michael Grubb, Plamen Nenov, Pian Shu, and Luis Zermeno.

Finally, I would like to thank my friends and family who have helped me overcome the challenges I have encountered over the past five years, especially Pian.

Contents

1 **Learning and Euler Equation Errors** **11**

1.1 Introduction 11

1.2 Background and Framework 15

1.2.1 Foundations 15

1.2.2 Estimation 17

1.2.3 Data 19

1.2.4 Pricing Error Puzzle 20

1.3 Learning and Estimation 22

1.3.1 Example of Pricing Errors Under Bayesian Learning 23

1.4 Model 25

1.4.1 Beliefs 27

1.4.2 Asset Prices 29

1.5 Simulations 31

1.6 Empirical 36

1.6.1 Methodology 37

1.6.2 Validation 40

1.6.3 Estimation 41

1.7 Conclusion 48

1.8 Appendix 50

1.8.1 Data Summary Statistics 50
1.8.2 Additional Simulations 50
1.8.3 Forecast Comparisons 50
1.8.4 Efficient Estimation 54
1.8.5 Proofs 55

2 High-Dimensional Impulse Response Function Matching Estimators 56

2.1 Introduction 56
2.2 Impulse Response Function Matching Estimators 60
2.2.1 The Common Approach 61
2.2.2 The Bias Corrected Approach 62
2.3 Consistency with Many Reduced Form Parameters 63
2.3.1 Setup 64
2.4 Bias Matters Asymptotically: An Example 67
2.4.1 Example 67
2.4.2 Simulation Results 70
2.5 Example: Altig et al (2011) 71
2.5.1 The Model 72
2.5.2 Estimation Methodology and Data 75
2.5.3 Simulations 76
2.5.4 First-Stage Bias vs. Identification 77
2.5.5 Simulation Results 80
2.5.6 Estimation Results 84
2.5.7 Evidence on Precision of Original Estimates 85
2.6 Conclusion 90
2.7 Appendix 91
2.7.1 Proofs 91

2.7.1.1 Proof of Theorem 1 (Positive Result) 91
2.7.1.2 Proof of Theorem 5 96
2.7.2 Tables and Figures 105

3 Carry Trade and Systemic Risk: Why are FX Options so Cheap? 109
3.1 Introduction 109
3.2 Carry Trade Basics 113
3.3 Risk-based Explanations 115
3.4 Data 117
3.5 Carry Trade and Systemic Risk: Pre-Crisis 118
3.5.1 Traditional Carry Trade Portfolios 118
3.5.2 Options Hedged Carry Trade 125
3.5.3 A New Puzzle 130
3.6 Carry Trade and Systemic Risk: Full Sample 132
3.7 Final Remarks 144
3.8 Appendix 146
3.8.1 Fama-French Risk Factors 146
3.8.1.1 Pre-Crisis 147
3.8.1.2 Full Sample 148

4 Reputation in Pay-as-You-Go Auctions 150
4.1 Introduction 150
4.2 Background and Data 153
4.2.1 The Basics of Pay-As-You-Go Auctions 153
4.2.2 Data 154
4.3 Baseline Model and Performance 158
4.3.1 Baseline Model 158

4.3.2 Model Performance on Auction Outcomes 161
4.3.3 Entertainment Values / Bid Costs 164
4.3.4 Risk Loving Behavior 165
4.4 Reputation 166
4.4.1 Descriptive Statistics 167
4.4.2 Empirical Results 170
4.4.3 Robustness 183
4.5 Conclusion 185
4.6 Appendix 187

Chapter 1

Learning and Euler Equation Errors

1.1 Introduction

The standard consumption-based asset pricing model with CRRA preferences provide a poor fit to real data on the returns of risky stocks and riskless bonds. The model produces very large pricing errors that cannot be explained with large values of risk-aversion or time discount rates, unlike the equity premium puzzle. Furthermore, the recent work of Lettau and Ludvigson (2009, hereafter LL) demonstrates that current leading consumption-based models cannot rationalize this *pricing error puzzle*.[1]

The primary contribution of this paper is to show that learning can explain the pricing error puzzle. Parameter uncertainty drives a wedge between agents' Euler equations and those that would emerge under full-information. This wedge causes the econometrician's moment conditions to fail in finite samples, which in turn manifests as pricing errors. First, I illustrate these results in a simple model, and prove that the theorem in LL does not extend to the case of learning. Second, I show that adding learning to an otherwise standard one factor model with CRRA preferences can generate quantitatively realistic pricing errors.

[1]This is shown specifically for the models of Bansal and Yaron (2004), Guvenen (2009), and Menzli et al. (2004), which is a multi-asset extension of Campbell and Cochrane (1999). Lettau and Ludvigson (2005) also show that departures from normal data generating processes produce only miniscule pricing errors.

Finally, I estimate a model with learning that makes minimal assumptions about the data generating process (DGP). After accounting for learning, the standard model with CRRA preferences produces pricing errors that are not statistically different from zero, despite having less volatile empirical moments. This paper is the first to propose a solution to the pricing error puzzle since LL.

This paper also contributes to a recent debate on whether models with learning alone can reasonably generate equity risk premia of the same magnitude that is observed in the data. Weitzman (2007) shows that even in a very simple model where consumption and dividends are equal, there exists some prior that can justify any observed average return to equities for a given sample size. On the other hand, Bakshi and Skoulakis (2010) argues that the prior required by this model to fit the historical data is unreasonable, implying implausible levels of structural uncertainty. Here, I consider learning in a less stylized model where dividends are not identical to consumption. Rather than assuming them, I construct "priors" by endowing agents with a dataset simulated from the true model at the time when they begin trading assets. I also assume that agent's posterior predictive distribution is normal rather than the thicker tailed student-t that is used in Weitzman (2007). This method is conservative in terms of how much structural uncertainty the agents possess. In simulations, I find that learning can generate an equity risk premium of roughly three quarters the size of that from the data. When estimating my empirical model, I find that learning typically reduces the estimated risk-aversion coefficient by a large margin, though not enough to fully explain the equity premium puzzle.

The work of LL provides a diagnosis for why models at the current frontier of consumption-based asset pricing have commonly been rejected in the data.[2] That is, these models have

[2]Fillat and Garduno (2005) reject the Campbell and Cochrane (1999) model with habits under three different hypotheses: complete markets, limited stock market participation, and incomplete markets. Other papers that empirically conclude against the Campbell and Cochrane (1999) model include Chen and Ludvigson (2009) and Tallarini and Zhang (2005) while Santos and Veronesi (2006) point out that the model still requires a risk aversion level of roughly 80 to match the data. Constantinides and Ghosh (2011) reject the Bansal and Yaron (2004) model, pointing to its over reliance on predictability in consumption growth

kernels that are too close, in a reduced form sense, to the benchmark kernel arising from CRRA preferences. More precisely, LL shows that if asset prices were generated from the models of Cambell and Cochrane (1999, hereafter CC), Bansal and Yaron (2004) (hereafter BY) or Guvenen (2009), an econometrician assuming CRRA preferences would observe no pricing errors for the benchmark model, although they would find implausible estimates for the preference parameters. In contrast, I show that a model with learning produces no pricing errors and is not rejected, but it can only partially explain the historical equity premium. Therefore, my results suggest that learning is likely to complement the mechanisms in these models well, though this is left for future work.

This project builds on past work in the literature on learning in consumption-based asset pricing models. Timmerman (1993, 1996) shows that when agents do not know the non-stochastic mean of iid dividend growth rates, asset returns are more volatile than otherwise. Brennan and Xia (2001) shows that if the mean of dividend growth rates is stochastic and unobservable but all non-stochastic parameters are known, the equity premium can be explained, but only with a risk aversion coefficient of 15 and a discount rate above 1. My model combines the elements of Timmerman (1993, 1996) and Brennan and Xia (2001) by assuming there is a stochastic, unobservable growth rate of dividends and consumption, and agents only have a finite amount of data from which to estimate the non-stochastic parameters. As discussed above, it is related to the work of Weitzman (2007) which constructs a model where some prior can explain any average equity premium. It is also related to Fuster, Laibson, and Mendel (2010) which argues that a model where agents use only a limited number of dependent variables in regressions can generate several key features of macroeconomic and asset price data. Finally, Adam, Marcet, and Nicolini (2008) shows that a simple model with self-referential learning can explain many of the key facts about asset prices, though their main results rely crucially on the assumption that dividends equal consumption.

---

as its main shortcoming, which echoes the criticism made by Beeler and Campbell (2009).

Other related work includes the literature on pessimistic beliefs as an explanation for the equity premium puzzle. Examples of papers which exogenously impose that agents are pessimistic relative to the historical data include Cecchetti, Lam and Mark (2000) and Abel (2002). Unlike those papers, I assume that agents use only in-sample observations to form their beliefs. Other papers take a robust control approach to endogenize pessimism by assuming that agents cannot form priors over models, and as a result, they act according to their worst case scenario model. Papers in that literature include Hansen, Sargent, and Tallarini (1999), Cagetti, Hansen, Sargent, and Williams (2002), Hansen, Sargent, and Wang (2002), and Anderson, Hansen, and Sargent (2003). Finally, Veronesi (2004) and Cogley and Sargent (2008b) combine exogenously imposed pessimism with learning, such that the pessimism loses relevance slowly enough to explain the size of the equity premium in a reasonably small sample. Unlike these papers, I do not assume that agents are excessively pessimistic relative to the data or that they prefer to act on their perceived worst-case scenario.

The remainder of the paper is organized as follows. Section 2 introduces the basic theoretical and empirical frameworks and describes the data and pricing error puzzle. Section 3 discusses a simple example to illustrate the impact of learning on pricing errors and proves that the main theorem in LL does not hold under learning. Section 4 describes the properties of a one factor model with learning. Section 5 presents simulations from the one factor model, showing that it can produce realistic pricing errors and a substantial equity premium. Section 6 presents and estimates an empirical model. Section 7 concludes.

1.2 Background and Framework

1.2.1 Foundations

To begin, I set out the basic framework and assumptions which the rest of the paper will build on. Suppose there is an infinitely-lived representative agent who consumes a single good. The agent's utility function is given by

$$U(C_t) = \frac{C_t^{1-\gamma_0} - 1}{1-\gamma_0}$$

where $\gamma > 0$ is the coefficient of relative risk aversion. Each period, indexed by t, the agent receives labor income, W_t, and has the option to invest in any combination of J assets, each with net supply normalized to 1. Asset prices are denoted by $P_{j,t}$, and dividends are denoted by $D_{j,t}$. The data generating process (DGP) for $(W_t, D_{1,t}, \ldots, D_{J,t})$, denoted Π_0, is exogenous and stationary.

The agent maximizes utility by solving

$$\max_{\{C_{t+k}\}_{k=0}^{\infty}} \sum_{k=0}^{\infty} \beta_0^k E_t^{\hat{\Pi}_t} \left(\frac{C_{t+k}^{1-\gamma_0} - 1}{1-\gamma_0} \right)$$

subject to the budget constraint

$$C_t + \sum_{j=1}^{J} P_{j,t} x_{j,t} \leq \sum_{j=1}^{J} D_{j,t} x_{j,t-1} + W_t$$

where $x_{j,t}$ is the amount that the agent chooses to invest in asset j at time t, $\beta \in (0,1)$ is the agent's time discount factor, and $\hat{\Pi}_t$ denotes the agent's beliefs about Π_0 given their information set as of time t. Importantly, I allow for $\hat{\Pi}_t \neq \Pi_0$.

The well-known necessary equilibrium conditions for this economy are the following set of Euler Equations and market clearing conditions:

$$1 = E_t^{\hat{\Pi}_t}\left[\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} R_{j,t+1}\right]\ \forall j, t$$

$$\begin{aligned} C_t &= \sum_{j=1}^{J} D_{j,t} + W_t\ \forall t \\ x_{j,t} &= 1\ \forall j, t. \end{aligned}$$

That is, asset prices are endogenous, and in equilibrium, they are given as the solution to the Euler Equations. Importantly, asset prices depend on the agent's beliefs $\hat{\Pi}_t$. Although Π_0 is directly defined as the distribution of $(W_t, D_{1,t}, \ldots, D_{J,t})$, the first market clearing condition implies that Π_0 also defines the distribution of C_t in equilibrium. In later sections, it will be more convenient to specify directly the DGPs for C_t and $D_{j,t}$, leaving the process for W_t defined implicitly.

The assumptions of a representative agent and exogenous labor income are for clarity of exposition and could be relaxed. In particular, the representative agent could be replaced by a continuum of agents with identical preferences and complete markets. Exogenous labor income could be replaced with perfectly competitive good producers such that the process for the sum of agents' marginal products of labor is given by W_t and the assumption that agents supply labor inelastically. Therefore, one can think of the model as the reduced form of a somewhat more general framework.

More generally, asset pricing models under complete markets and no arbitrage take the form

$$1 = E_t^{\Pi_0}\left[M_{t+1} R_{j,t+1}\right]$$

where M_{t+1} is known as the pricing kernel. In consumption-based models, like the one employed herein, the kernel is equal to the intertemporal ratio of the agent's marginal utilities.

1.2.2 Estimation

Following the seminal work of Hansen and Singleton (1982), GMM is the standard method for estimating the preference parameters, denoted $\phi \equiv (\beta, \gamma)$, for the model in the preceding section. In practice, the econometrician observes aggregate consumption and asset returns for a total of T periods. He selects K assets of interest, and computes parameter estimates by minimizing a measure of the distance between observed asset prices and those implied by the model. More specifically, parameter estimates are computed as

$$\begin{aligned} \hat{\phi}_T &= \arg\min_{\phi} Q_T(\phi) \\ Q_T(\phi) &\equiv \mathbf{g}_T'(\phi)\,\mathbf{W}_T\mathbf{g}_T(\phi) \end{aligned}$$

where $\mathbf{W}_T$ is any positive definite matrix, $\mathbf{g}_T(\phi) = [g_{1,T}, \ldots, g_{K,T}]'$, and moments $g_{j,T}$ are computed as

$$\begin{aligned} g_{j,T}(\phi) &= \frac{1}{T}\sum_{t=1}^{T} u_{j,t}(\phi) \\ u_{j,t+1}(\phi) &= \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{j,t+1} - 1. \end{aligned}$$

Here, $u_{j,t}$ are commonly referred to as the pricing errors from the model, and the procedure selects parameter estimates to minimize their (weighted) Euclidean distance from zero.

The crucial assumption for the consistency of GMM is that

$$E^{\Pi_0}\left[u_{j,t}(\phi_0)\right] = 0$$

where $\phi_0 = (\beta_0, \gamma_0)$ refers to the agent's true preference parameters. In the most standard benchmark consumption-based model, the agent is assumed to have CRRA utility and full information about the DGP (i.e., $\hat{\Pi}_t = \Pi_0$). That is, the agent knows with perfect accuracy the true distribution of dividends and output. Under these assumptions, one can rewrite the agent's Euler Equation as

$$1 = E_t^{\Pi_0}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{j,t+1}\right]$$

which, by the law of iterated expectations, implies

$$E^{\Pi_0}\left[u_{j,t}\left(\phi_0\right)\right] = 0.$$

Thus, the necessary condition for the consistency of GMM is satisfied, and it can be used to estimate the benchmark model.

Two choices for $\mathbf{W}_T$ are common. The first and the simplest is the identity matrix, $\mathbf{W}_T = I_K$. For this choice, GMM places equal weights on all chosen assets and seeks to minimize their squared pricing errors. The appeal of identity weighting is that it minimizes the model's pricing errors for the chosen assets directly. Since the included assets are generally chosen based on economically interesting criteria, setting $\mathbf{W}_T = I_K$ allows one to assess the model's fit on the set of economically interesting moments. For any other weighting matrix, the procedure minimizes linear combinations of the economically interesting moments, which can make the results harder to interpret. When the identity matrix is employed, one can compute an economically meaningful measure of fit as

$$RMSE_T \equiv \sqrt{\frac{1}{K}\sum_{k=1}^{K} g_{k,T}^2\left(\hat{\phi}_T\right)}$$

This measures the size of the average pricing errors $g_{k,T}$ from the model. In order to better guage its magnitude, $RMSE_T$ can be compared with

$$\begin{aligned} RMSR_T &\equiv \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\bar{R}_{k,T} - 1\right)^2} \\ \bar{R}_{k,T} &= \frac{1}{T}\sum_{t=1}^{T} R_{k,t}. \end{aligned}$$

Following LL, I use $\frac{RMSE_T}{RMSR_T}$ as a goodness-of-fit measure for GMM with identity weighting. The fit is better when this measure is closer to zero.

The second common choice of weighting matrix is $\mathbf{W}_T = \hat{\Sigma}_T^{-1}$ where $\hat{\Sigma}_T$ is a consistent estimator of $E\left[\mathbf{g}_T(\phi_0)\,\mathbf{g}_T(\phi_0)'\right]$. The appeal of this second choice is that it maximizes the efficiency of the estimator. It also allows for easy testing of the model. Let $Q_T^*(\phi) \equiv \mathbf{g}_T'(\phi)\,\hat{\Sigma}_T^{-1}\mathbf{g}_T(\phi)$, then $TQ_T^*\left(\hat{\phi}_T\right) \rightarrow_d \chi^2_{K-2}$, as long as $K > 2$. Comparing the values of $TQ_T^*\left(\hat{\phi}_T\right)$ to the critical values from χ^2_{K-2} is referred to as a test of overidentification. If $K = 2$, the procedure is exactly identified, and typically $Q_T\left(\hat{\phi}_T\right) = 0$, since estimation is equivalent to solving a system of two equations and two unknowns (the first order conditions for the minimization), which usually has an exact solution, even if the system is non-linear.

1.2.3 Data

For my empirical results, I use quarterly data from the United States for the period starting on the first quarter of 1960 and ending on the first quarter of 2008. I set aside the first 12 quarters of this data to initialize the learning model, as discussed in later sections, and I only use the remaining 181 quarters of data for estimation. Following LL, I use data on 8 different assets, including a market return (measured by the CRSP value weighted index return),[3] a risk-free return (measured by the three-month T-bill rate),[4] and the returns of 6 portfolios of equities sorted on size and book to market from Kenneth French's Dartmouth website.[5] Fama and French (1992) argue that this set of 8 assets provides a good representation of the cross-section of US equity returns. I measure C_t as real per capita consumption of nondurables and services, excluding shoes and clothing, as in LL.[6] Returns are converted

[3] The source for the CRSP value weighted return is the Center for Research in Security Prices (CRSP).

[4] Three-month T-bill rates are taken from H.15 Release-Federal Reserve Board of Governors.

[5] The website is located at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[6] The source for my consumption data is the U.S. Department of Commerce, Bureau of Economic Analysis (BEA). It is seasonally adjusted at annual rates. Population is computed by dividing real total disposable income by real per capita disposable income, both of which are also from BEA.

into real terms using the price deflator for the chosen consumption series.[7] In the appendix, I provide summary statistics for my data.

1.2.4 Pricing Error Puzzle

Having introduced the basic notation and estimation methodology, I turn to the pricing error puzzle. Table 1 shows the results of estimating the benchmark model with CRRA preferences and $\hat{\Pi}_t = \Pi_0$. The first line corresponds to the exactly identified case where the included assets are the market and risk-free returns. The results confirm that the model provides a poor fit to the data, failing along two important dimensions. First, the estimates of the preference parameters are well beyond reason. As discussed in Kocherlakota (1996), the consensus is that a reasonable value of γ for these preferences would be less than 10. Meanwhile, it is equally, if not even harder, to believe that $\beta > 1$, which would imply that people prefer consumption in the future over the present. Secondly, even for the implausible parameter estimates, the pricing errors are large relative to the underlying returns. The goodness-of-fit measure, $\frac{RMSE}{RMSR}$, shows that the pricing errors are over 40% of the size of the returns being matched. The second row of the table shows the results for the overidentified system that adds the six Fama-French portfolios to the market and risk-free returns. Qualitatively the results remain the same. The values of the preference parameters are implausible, and the pricing errors remain large. The final column shows the p-value from an overidentification test, and not surprsingly, the model is strongly rejected.

Table 1: Benchmark Estimates

Assets	$\hat{\gamma}$	$\hat{\beta}$	RMSE	RMSR	$\frac{RMSE}{RMSR}$	p-value
R_m, R_f	111	1.59	2.12%	5.12%	0.41	-
$R_m, R_f, R_1 - R_6$	92	1.48	3.33%	9.18%	0.35	5.67e-8

Notes: The p-value is for the overidentification test with heteroskedasticity and autocorrelation robust (Newey-West) covariance matrix.

[7] The source is BEA.

Over the past decade, extensions of the baseline model have been designed to fit key aggregate statistics of observed equity returns, such as the average historical equity premium. As mentioned previously, current leading models of US equities include CC, BY, and Guvenen (2009). The first two posit that CRRA utility is too restrictive, and instead propose more general and complex utility functions. Meanwhile the primary story in Guvenen (2009) is that not all individuals are stockholders, and therefore the consumption data used in the test shown in Table 1 is mismeasured. These stories are each plausible and relevant, but LL show that they all fail in one important regard. In particular, for data from any of these models, LL show

$$\exists \tilde{M}_{t+1} = \tilde{\beta}\left(\frac{C_{t+1}}{C_t}\right)^{-\tilde{\gamma}}, \text{ such that } E^{\Pi_0}\left[\tilde{M}_{t+1}R_{j,t+1}\right] = 1$$

This proves that, even if preferences are misspecified or consumption is mismeasured along the lines suggested by these papers, the moment condition used in GMM is still satisfied for some pseudo-kernel of the CRRA form. Therefore, the benchmark model should produce negligible pricing errors and not be rejected by a test of overidentification for some parameter values. To see why, notice that in the exactly identified case GMM is solving the system of two equations and two unknowns

$$\frac{1}{T}\sum_{t=1}^{T}\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{t+1} = 1$$

where R_{t+1} is the 2x1 vector of returns. Of course, the use of the sample average as opposed to the true $E^{\Pi_0}[.]$ operator in practice, opens the possibility that these pricing errors might not literally be equal to zero, but LL provide strong simulation evidence that this is not the case, finding that the pricing errors from these models in realistic sample sizes are negligible.

1.3 Learning and Estimation

In this section, I show how the presence of learning affects the properties of the GMM estimation procedure described above, and in particular, how it could produce pricing errors much like those seen in the data. First of all, it is easy to see that arbitrary deviations from $\hat{\Pi}_t = \Pi_0$ will lead to violations of the econometrician's moment conditions. That is,

$$E^{\Pi_0}\left[E_t^{\hat{\Pi}_t}\left[\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} R_{j,t+1} - 1\right]\right] \neq E^{\Pi_0}\left[\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} R_{j,t+1} - 1\right]$$

$$\Rightarrow \quad E^{\Pi_0}\left[u_{j,t}\left(\phi_0\right)\right] \neq 0$$

because the law of iterated expectations cannot be applied when the two expectations are taken under different measures. One may wonder, however, whether for reasonable specifications of $\hat{\Pi}_t$, the law of iterated expectations would still apply. In particular, will the econometrician's moment conditions be satisfied if the agent is a Bayesian learner? The answer turns out to be no. In fact, it turns out that the even if the agent is a Bayesian learner the moment conditions will not generally be satisfied *for any choice of* ϕ.

It is tempting to argue that the only difference between Bayesian learners and fully informed agents is that the Bayesian agents have a coarser information set which excludes the parameters. If this were the case then one could apply the law of iterated expectations to show that the econometrician's moment conditions still hold. However, if the true distribution of fundamentals is defined up to a parameter, θ, one cannot compute an expectation under $\Pi_0(\theta)$ without being given the value of θ, as it is part of the definition of Π_0. As a result, Bayesian agents' must specify priors which distort their subjective measure away from Π_0. The following simple example shows that the wedge between a Bayesian's subjective measure and the true measure can lead to pricing errors.

1.3.1 Example of Pricing Errors Under Bayesian Learning

In addition to the framework outlined above, assume that

$$\begin{pmatrix} D_t \\ \Delta c_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{d,0} \\ \mu_{c,0} \end{pmatrix}, I_2\right)$$

where $c_t = \log C_t$. Assume also that the agent observes the entire history of D_t and C_t, but they do not know the values $\mu_{d,0}$ or $\mu_{c,0}$. Instead, the agent is Bayesian with prior

$$\begin{pmatrix} \mu_d \\ \mu_c \end{pmatrix} \sim N\left(\begin{pmatrix} z_d \\ z_c \end{pmatrix}, \sigma^2 I_2\right).$$

Standard calculations show that the agent's posterior is given by

$$\begin{pmatrix} \mu_{d,t|t} \\ \mu_{c,t|t} \end{pmatrix} \sim N\left(\left(1+\sigma^2 t\right)^{-1}\begin{pmatrix} z_d + \sigma^2 \sum_{j=1}^{t} D_j \\ z_c + \sigma^2\left(c_t - c_1\right) \end{pmatrix}, \sigma^2\left(1+\sigma^2 t\right)^{-1} I_2\right)$$

Consider two assets, both one period bonds. The first bond is risk-free, paying 1 in period $t+1$ with price $P_{f,t}$, and the second is risky paying D_{t+1} with price $P_{1,t}$. Slightly rewriting the agent's Euler Equations from before, the two bonds' prices must satisfy

$$\begin{aligned} P_{f,t} &= \beta_0 E_t^{\hat{\Pi}_t}\left[\exp\left(-\gamma_0 \Delta c_{t+1}\right)\right] \\ P_{1,t} &= \beta_0 E_t^{\hat{\Pi}_t}\left[\exp\left(-\gamma_0 \Delta c_{t+1}\right) D_{t+1}\right]. \end{aligned}$$

Finally, consider what happens if an econometrician uses the standard GMM procedure, as outlined in the previous section, to estimate the parameters (β_0, γ_0). The following theorem shows that the LL result does not hold for this example, and therefore is not generally true under learning. The theorem shows the necessary and sufficient condition for the econometrician's moment condition to have a solution for some values of the preference parameters. Let $R_t = [R_{f,t}, R_{1,t}]'$ be the 2x1 vector of returns for the two bonds.

Theorem 1. *$\exists \tilde{M}_{t+1} = \tilde{\beta}\left(\frac{C_{t+1}}{C_t}\right)^{-\tilde{\gamma}}$ such that $E^{\Pi_0}\left[\tilde{M}_{t+1}R_{t+1}\right] = 1$*
if and only if
$\mu_{d,0}\left(1+\sigma^2 t\right)E^{\Pi_0}\left[\left(z_d + \sigma^2\sum_{j=1}^{t} D_j\right)^{-1}\right] = 1.$

Since the condition in Theorem 1 is not typically satisfied, I find that even in this simple example with Bayesian learning, the econometrician's moment conditions will not generally hold for *any* values of the preference parameters. As a special case, the following corrollary shows that the moment conditions are satisfied when the agent has full knowledge of the DGP.

Corollary 2. *If $\hat{\Pi}_t = \Pi_0$, $\exists \tilde{M}_{t+1} = \tilde{\beta}\left(\frac{C_{t+1}}{C_t}\right)^{-\tilde{\gamma}}$ such that $E^{\Pi_0}\left[\tilde{M}_{t+1}R_{j,t+1}\right] = 1$.*

This follows from the observation that $\hat{\Pi}_t = \Pi_0$ is equivalent to the set of conditions: $\sigma = 0$, $z_d = \mu_{d,0}$, and $z_c = \mu_{c,0}$. Plugging those into the condition from the theorem yields the result. Hence, if the agent has perfect information about the DGP, the econometrician's moment condition will be satisfied, thereby recovering the result in LL. If, however, the agent has some uncertainty about the parameters, captured by $\sigma > 0$, the econometrician's moment condition will typically be misspecified.

The purpose of the theorem is to shed light on the likelihood that we observe $RMSE_T = 0$ in practice. To see the connection between the result and the question of interest, first notice that $RMSE_T = 0$ if and only if $\frac{1}{T}\sum_{t=1}^{T}\tilde{M}_{t+1}R_{t+1} = 1$ for some $\tilde{M}_{t+1} = \tilde{\beta}\left(\frac{C_{t+1}}{C_t}\right)^{-\tilde{\gamma}}$. A natural question to ask is whether or not this condition will be satisfied, at least on average, for a sample of size T. Thus, the ideal theorem would make a statement about when $\frac{1}{T}\sum_{t=1}^{T}E^{\Pi_0}\left[\tilde{M}_{t+1}R_{t+1}\right] = 1$ holds. Although the stated theorem excludes the exterior sample average in this expression, this difference is immaterial because given that $E^{\Pi_0}\left[\tilde{M}_{t+1}R_{t+1}\right] \neq$

1, it will not generally be true that $\frac{1}{T}\sum_{t=1}^{T} E^{\Pi_0}\left[\tilde{M}_{t+1}R_{t+1}\right] = 1$. As a result, we can interpret the result of the theorem as saying that under learning, for a sample of size T, it is more common to have $RMSE_T > 0$ than under full information.

1.4 Model

Although the preceding theoretical result shows that learning leads to pricing errors in general, it does not imply that the pricing errors induced by learning are large enough to match the data. In order to quantify this effect, I construct a richer model. Using this model, I also consider the quantitative implications of learning for risk premia.

Keeping the basic framework of a representative agent with CRRA preferences, I introduce new assumptions on Π_0 and the agent's beliefs. Specifically, I assume Π_0 is given by the following one-factor model

$$x_t = \rho x_{t-1} + \sigma_x e_{x,t} \tag{1.4.1}$$

$$\Delta c_t = \mu_c + x_t + \sigma_c e_{c,t} \tag{1.4.2}$$

$$\Delta d_t = \mu_d + \psi x_t + \sigma_d e_{d,t} \tag{1.4.3}$$

where c_t, d_t denote the log of consumption and dividends respectively, $e_{i,t}$ is iid standard normal for all i, and $E\left(e_{i,t}e_{j,t}\right) = 0$ for all i,j. Let $\theta \equiv (\rho, \psi, \mu_c, \mu_d, \sigma_x, \sigma_c, \sigma_d)$ be the vector of DGP parameters.

This choice of DGP is common in the literature. BY study the case where the state, x_t, and all parameters are observable. In that case, one can generate realistic asset returns with Epstein-Zin utility, a coefficient of relative risk aversion of 10, and a persistent state ($\rho = 0.96$). Brennan and Xia (2001) study the case where the state, x_t, is unobservable and does not enter the consumption equation. Under those circumstances and when agents have CRRA preferences, a relative risk aversion coefficient of 15 and negative discount factor

$(\beta > 1)$ are needed to match the equity premium and risk-free rate in the data. My primary goal will be to consider the realistic case where the state x_t is unobservable and parameters are unknown, but first I will briefly describe the impact that parameter uncertainty has on risk-premia for a simpler case where only ρ is unknown.

As before, consider for now just risk-free and risky one period bonds. Let $r_t^f = -\log(P_{f,t})$ and $r_{t+1}^1 = d_{t+1} - \log(P_{1,t})$ be the log risk-free and risky returns, respectively. The following theorem characterizes how the risk-free rate and risk premium are affected by parameter uncertainty.

Theorem 3. *If x_t and all parameters are observable,*

$$r_t^f = \gamma\mu_c + \gamma\rho x_t - \log(\beta) - \frac{\gamma^2}{2}\left(\sigma_x^2 + \sigma_g^2\right)$$

$$r_{t+1}^1 - r_t^f = \psi\sigma_x\varepsilon_{x,t+1} + \sigma_d e_{d,t+1} + \psi\left(\gamma - \frac{\psi}{2}\right)\sigma_x^2 - \frac{1}{2}\sigma_d^2.$$

If x_t and all parameters are observable except ρ and the agent has a normal posterior over ρ, then

$$r_t^f = \gamma\mu_c + \gamma\hat{\rho}_t x_t - \log(\beta) - \frac{\gamma^2}{2}\left(\sigma_x^2 + \sigma_g^2 + x_t^2 R_{t|t}\right)$$

$$r_{t+1}^1 - r_t^f = (\rho - \hat{\rho}_t)\,\psi x_t + \psi\sigma_x e_{x,t+1} + \sigma_d e_{d,t+1} + \psi\left(\gamma - \frac{\psi}{2}\right)\left(\sigma_x^2 + x_t^2 R_{t|t}\right) - \frac{1}{2}\sigma_d^2$$

where $\hat{\rho}_t$ and $R_{t|t}$ are the mean and variance, respectively, of the agent's posterior for ρ.

Theorem 2, shows three specific noteworthy effects of uncertainty about ρ as compared with the full information case: (1) the risk free rate is lower; (2) the expected risk premium is larger whenever $\gamma > \frac{\psi}{2}$; and (3) average returns are time-varying, though the agent is not able to take advantage of this due to their parameter uncertainty. More generally, the deviations from full-information returns can be characterized as coming from two sources. The first is the bias in the agent's beliefs, which in this example corresponds to the case when $E^{\Pi_0}(\hat{\rho}_t) \neq \rho$. The second comes from the degree of the agent's uncertainty about parameters, which he views similarly to standard sources of uncertainty. This occurs in the example when $R_{t|t} > 0$ and naturally provides a direct additive term into risk premia and risk-free rates. As long as the agent is able to consistently estimate θ, both of these factors will vanish asymptotically. However, the impact of parameter uncertainty may last a long time, or as in Weitzman (2007), if there are periodic structural breaks in θ, the agent can never fully learn its true value. When the state and more parameters are unobservable, the expression is not available in closed form, so next I explore that case numerically.

1.4.1 Beliefs

To close the model, I turn to a description of how the agent forms beliefs about the DGP parameters from the observed data. The most natural specification for expectations when the DGP is imperfectly known is Bayesian learning, but this is not straightforward to do for this model. The challenge arises because the standard conjugate prior for the volatility parameters implies that the posterior predictive distribution is student-t. Since the moment generating function of a student-t distribution does not exist, the prices of assets do not exist and expected utility need not exist. In this paper, I take a different approach. I assume that the agent estimates DGP parameters by maximum likelihood, and uses the estimated asymptotic distribution to form expectations. This assumption may be thought of as saying that the agent forms expectations just as if he were a frequentist econometrician.

To be more precise, the agent observes $y_t = (C_t, D_t)$ each period. Let $\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_0)$ denote the history of data through time t. The agent has full information about Π_0, except he does not know the value of θ, and does not observe the state, x_t. That is, he knows perfectly the distribution $\Pi_0(\mathcal{Y}_t|\theta, x_t)$. Also he can compute the distribution $\Pi_0(\mathcal{Y}_t|\theta)$ via the Kalman filter.[8] As such, agents can perform maximum likelihood estimation of θ as

$$\hat{\theta}_T = \arg\max_{\theta} \sum_{t=1}^{T} \log \Pi_0 \left(y_t | \theta, \mathcal{Y}_{t-1} \right).$$

These estimates $\hat{\theta}_T$ describe the agent's best guess about the true value of the DGP parameters. In order to assess his own uncertainty about his estimates, the agent does just as a frequentist econometrician would, by turning to MLE asymptotic theory, which states that

$$\sqrt{T} \left(\hat{\theta}_T - \theta_0 \right) \to_d N \left(0, H \right)$$

where

$$H = -\frac{1}{T} E^{\Pi_0} \left[\sum_{t=1}^{T} \left. \frac{\partial^2 \log \Pi_0 \left(y_t | \theta, \mathcal{Y}_{t-1} \right)}{\partial \theta \partial \theta'} \right|_{\theta = \theta_0} \right].$$

In practice, the agent does not observe H, just as he does not observe θ_0. Instead, the agent forms his beliefs as

$$\hat{\Pi}_t \left(\theta | \mathcal{Y}_t \right) = N \left(\hat{\theta}_t, \frac{1}{t} \hat{H}_t \right)$$

where

$$\hat{H}_t = -\frac{1}{t} \sum_{t=1}^{T} \left. \frac{\partial^2 \log \Pi_0 \left(y_t | \theta, \mathcal{Y}_{t-1} \right)}{\partial \theta \partial \theta'} \right|_{\theta = \hat{\theta}_t}.$$

Although there are reasons for agents to be Bayesian that are rooted in utility maximization as pointed out by Savage (1954), the non-Bayesian specification employed herein has its advantages as well. As pointed out earlier, prior conjugacy in this case leads to an infinite equity premium and no equilibrium. This problem could in principle be solved, as in Weitzman (2007), by introducing new hyperparameters which define the boundaries of the support of the agent's prior over the volatility parameters. Although such an approach preserves Bayesian learning, it also leads to a model that is difficult to evaluate empirically.

[8]The exact implementation of the Kalman filter for this model is standard and can be found in Hamilton (1994).

The choice of prior boundary values is sufficiently abstract that it is hard to justify particular numbers, and Weitzman (2007) showed that asset price dynamics are extremely sensitive to these choices. In contrast, the approach I take does not require the subjective specification of priors or hyperparameters.

1.4.2 Asset Prices

For tractibility, until now I have introduced only one risky asset, a one period bond, but in order to assess the model's performance on equities, I introduce a market return with capital gains. Assume that there is a third asset that can be traded which has an infinite horizon and pays D_t in each period t. I will refer to this as the market asset, and its price must satisfy

$$P_t^m = E_t^{\hat{\Pi}_t}\left[\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}\left(P_{t+1}^m + D_{t+1}\right)\right],$$

which can be expanded as

$$P_t^m = E_t^{\hat{\Pi}_t}\left[\beta_0\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0} D_{t+1}\right] + E_t^{\hat{\Pi}_t}\left[\beta_0^2 E_{t+1}^{\hat{\Pi}_{t+1}}\left[\left(\frac{C_{t+2}}{C_t}\right)^{-\gamma_0}\left(P_{t+2}^m + D_{t+2}\right)\right]\right].$$

In order to simplify this expression, I assume that the agent believes his beliefs will be consistent over time, so that $E_t^{\hat{\Pi}_t}\left[E_{t+1}^{\hat{\Pi}_{t+1}}[x]\right] = E_t^{\hat{\Pi}_t}[x]$. If the agent were truly Bayesian or had full information about the DGP, then this assumption would hold true. However, given that the agent is non-Bayesian, this assumption makes the agent's beliefs slightly inconsistent over time. That is, the agent pretends as though he will be using Bayes' rule to incorporate the data from next period into his beliefs, even though he knows that he will not literally do that. Although imperfect, this assumption is necessary for computational tractability of market prices. It is also much weaker than the typical assumption in the theoretical literature on learning and asset prices. The standard assumption is that when making decisions agents treat estimated parameters, $\hat{\theta}_T$, as perfectly known. This would be equivalent to setting $\hat{H}_T = 0$.[9] Here, I strive to relax that assumption as far as possible while maintaining tractibility.

Given the assumptions on the agent's beliefs, one can expand the previous expression recursively to obtain

$$\frac{P_t^m}{D_t} = \sum_{j=1}^{\infty} \beta^j \hat{E}_t \left[\exp\left(\sum_{k=1}^{j} \left(\Delta d_{t+k} - \gamma \Delta c_{t+k} \right) \right) \right],$$

which can be written as

$$\frac{P_t^m}{D_t} = \sum_{j=1}^{\infty} \beta^j E_t^{\hat{\Pi}_t} \left[\exp\left(f\left(j, \theta, \mathcal{Y}_t\right)\right)\right] \tag{1.4.4}$$

where

$$\begin{aligned}
f\left(j, \theta, \mathcal{Y}_t\right) &= j\left(\mu_d - \gamma \mu_c\right) + \left(\psi - \gamma\right) \rho \left(\frac{1-\rho^j}{1-\rho}\right) x_{t|t}\left(\theta\right) \\
&+ \frac{1}{2}\left[\left(\psi - \gamma\right) \rho \left(\frac{1-\rho^j}{1-\rho}\right)\right]^2 P_{t|t}\left(\theta\right) \\
&+ \frac{1}{2}\left(j\left(\sigma_d^2 + \gamma^2 \sigma_c^2\right) + \left(\psi - \gamma\right)^2 \sigma_x^2 \sum_{k=1}^{j} \left(\frac{1-\rho^{j-k+1}}{1-\rho}\right)^2 \right) \\
x_{t|t}\left(\theta\right) &= E^{\hat{\Pi}_t}\left[x_t | \mathcal{Y}_t, \theta\right] \\
P_{t|t}\left(\theta\right) &= E^{\hat{\Pi}_t}\left[\left(x_{t|t} - x_t\right)^2 | \mathcal{Y}_t, \theta\right].
\end{aligned}$$

The terms $x_{t|t}(\theta)$ and $P_{t|t}(\theta)$ are deterministic functions of the history of data and parameters that can be computed via the Kalman filter. Thus, the only random variables from the agent's point of view in $f(j, \theta, \mathcal{Y}_t)$ are the parameters. The final step to computing the market price involves taking expectations over the agent's beliefs about θ. Unfortunately, this expression cannot be simplified further analytically, but it can be computed by simulation, as I discuss in the following section.

[9] See, for example, Timmerman (1993, 1996) or Cogley and Sargent (2008b).

1.5 Simulations

In this section, I show that parameter uncertainty, even in a relatively simple model, can explain the pricing errors of the benchmark model and generate a sizeable equity risk premium. These results contrast past work that has used the same DGP and preferences but endowed agents with full information. Specifically, BY shows that this one factor DGP with full-information and CRRA preferences cannot produce anything near realistic risk premia, and LL shows that it also cannot generate realistic pricing errors.

I begin by estimating the parameters of the DGP by maximum likelihood on my full sample of quarterly data. The data on dividend growth is obtained by combining the CRSP value weighted returns with the corresponding ex-dividend return in a standard way. The estimated parameters along with standard errors are reported in Table 2. Several features are worth mentioning. First, BY shows that the value of ρ matters for asset pricing predictions. In particular, it argues that under full information, $\rho \approx 1$ is important for generating a realistic equity premium, even with recursive preferences. However, Table 2 reveals that this is not a realistic feature of the data, a fact that has been pointed out elsewhere by Beeler and Campbell (2009) and Constantinides and Ghosh (2011). Second, the dynamics of consumption and dividend growth are substantially different. Consumption growth has a higher mean and lower volatility than dividend growth. This suggests that the one factor model employed herein is a meaningful extension of recent models on asset prices under learning in Weitzman (2007) and Bakshi and Skoulakis (2010) which equate these two processes. Finally, one can see that there are non-negligible standard errors around certain key parameter estimates, indicating a potential role for parameter uncertainty. In particular, a confidence interval for the mean of dividend growth contains a wide range of values, including zero, which necessarily have very different implications for the equity premium.

Table 2: DGP Parameter Estimates

	ψ	ρ	μ_c	μ_d	σ_c	σ_d	σ_x
Estimate	1.34	0.76	0.005	0.003	0.003	0.11	0.002
Std. Error	1.98	0.06	0.0004	0.006	0.0002	0.004	0.0002

Notes: Parameters are estimated by maximum likelihood.

Given values for the DGP parameters, I simulate data from the model and compute asset prices. The simulations are computed in the following steps for each t:

1. Simulate $\Delta c_t, \Delta d_t, x_t$.
2. Compute $\hat{\theta}_t$ and $\hat{H}_t$ by maximum likelihood.[10]
3. Simulate observations from $\theta_{(b)} \sim N\left(\hat{\theta}_t, \frac{1}{t}\hat{H}_t\right)$ for $b = 1, \ldots, 100$.[11]
4. For each b, compute $P_{f,t}^{(b)}$ and $P_{m,t}^{(b)}$ as if $\theta_{(b)}$ were known to be true.
5. Compute $P_{f,t} = \frac{1}{100}\sum_{b=1}^{100} P_{f,t}^{(b)}$ and $P_{m,t} = \frac{1}{100}\sum_{b=1}^{100} P_{m,t}^{(b)}$.
6. Compute returns $r_{f,t} = \frac{1}{P_{f,t-1}}$ and $r_{m,t} = \frac{P_{m,t}+D_t}{P_{m,t-1}}$.

Generating a realistic risk premium in this model requires a low value of risk aversion, γ, because for CRRA preferences, this is also equal to the inverse of the intertemporal elasticity of substitution. Unfortunately, however, the price of the market asset for low values of γ does not exist for this DGP under either full information or learning. To overcome this problem, I modify the parameters of the DGP so that market prices will exist. The only change I make is to lower σ_d from its estimated value of 0.11 to 0.04. Although this modified DGP is no longer a good fit to the volatility of dividends, the change should be relatively innocuous, as it has virtually no impact on asset prices for the full-information model. For example, if I use $\gamma = 5$, the full information average risk-premium changes from 0.03% with the estimated

[10] I estimate the parameters for every other t to decrease computation times.

[11] Draws $\theta_{(b)}$ which lead to non-existent asset prices are discarded (e.g., those where $\rho > 1$).

σ_d to -0.01% with the lower value. This is consistent with the results of BY which shows that e_d is an unpriced risk factor, since it is uncorrelated with the pricing kernel. Moreover, this modified dividend process is still more realistic than the one considered by Weitzman (2007) and Bakshi and Skoulakis (2010). As a robustness check, I include simulation results for the maximum likelihood estimated parameters and $\gamma = 5$ in Table A2 of the appendix. Those results confirm that learning generates larger pricing errors and risk premia than the full information model.

With the lower value of $\sigma_d = 0.04$, I am able to compute market prices for lower values of risk-aversion. Table 3 reports the results when setting $\gamma = 0.5$ and $\beta = 0.997$ for 400 simulated samples each of length 181 observations, the same sample size that I use for estimation. Since it is reasonable to assume that people had data going back historically further than my sample, I also give the agent an initial 50 quarters of data on consumption and dividends. That is, I simulate 231 quarters of data for $\Delta c_t, \Delta d_t, x_t$, and compute asset prices for the latter 181 of those observations. This is similar to endowing the agent with a prior, but constructing it in this fashion ensures the "prior" is reasonable. Constructing initial beliefs in this way also avoids introducing more degrees of freedom into the analysis, and thereby alleviates concerns about robustness along that particular dimension of the model.

Table 3 shows that the model with learning provides a broad improvement over the full-information model. The third row contains key statistics from the data, while the first two rows show the averages across simulated samples of these statistics for the models with learning and full-information. The last two columns report the 90% and 95% quantiles of the distribution of the full information model's pricing errors. That is, I perform the same exercise as in Table 1 on each simulated sample from the two models, and compute quantiles across simulations. These two quantiles were chosen as they correspond to typically used significance levels in classical hypothesis testing. Comparing the first and second rows, learning generates large increases in both risk premia and pricing errors relative to the full-information model. Learning is able to produce pricing errors of the same magnitude and an equity premium that is roughly three quarters of what is observed in the data, whereas the full information model is not close on either of these key dimensions.

Table 3: Simulation Results

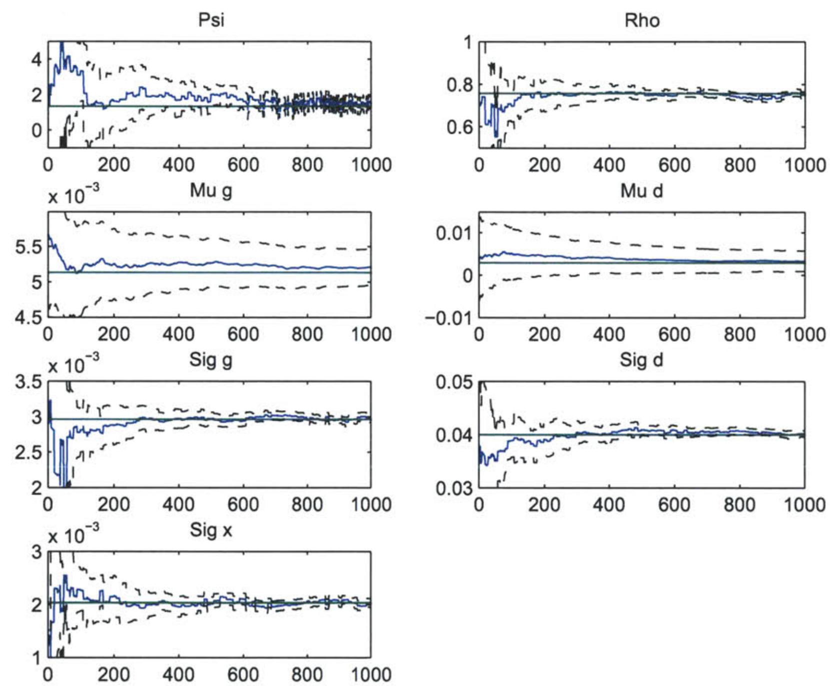
	$E[r_m - r_f]$	$\sigma[r_m]$	$E[r_f]$	$\sigma[r_f]$	$Q\left(\frac{RMSE}{RMSR}, 90\%\right)$	$Q\left(\frac{RMSE}{RMSR}, 95\%\right)$
Learning	1.08%	8.35%	0.56%	0.08%	0.41	0.47
Full Info.	0.00%	4.11%	0.56%	0.12%	0.12	0.24
Data	1.44%	8.32%	0.34%	0.57%	0.41	0.41

Notes: Preference parameters are $\beta = 0.997$, $\gamma = 0.5$. DGP parameters are set at the maximum likelihood estimates from Table 2, except $\sigma_d = 0.04$. $Q(X, y)$ denotes the y-th quantile of X.

The effects of learning comes from two potential sources. The first is estimation bias, the systematic difference between estimated and true parameter values for the relevant sample sizes. The second source is the degree of parameter uncertainty, which I measure by the width of estimated confidence bands. Since the agent's beliefs about parameters is always described by a normal distribution, focusing on the width of the confidence interval is sufficient for his uncertainty, except that it omits how beliefs about different parameters are correlated. The correlation of beliefs across parameters is, however, taken into consideration when computing asset prices.

In order to provide an illustration of these two effects, Figure 1 shows the average paths of estimated parameters for long samples of simulated data along with their corresponding average 95% confidence bands. Here, I use 10 simulations of 1,050 observations, and report estimates for the final 1,000 observations of each. One can see that in small samples, of similar sizes to what is encountered in the data, both bias and parameter uncertainty are non-negligible. For example, the confidence interval for μ_d includes zero for roughly the first 700 quarters. Although the bias disappears for most parameters after about 250 quarters, the uncertainty remains for considerably longer in many cases. While the model with learning converges to a model where the agent is a formally Bayesian about the state, Figure 1 illustrates that this convergence takes a substantial amount of time.

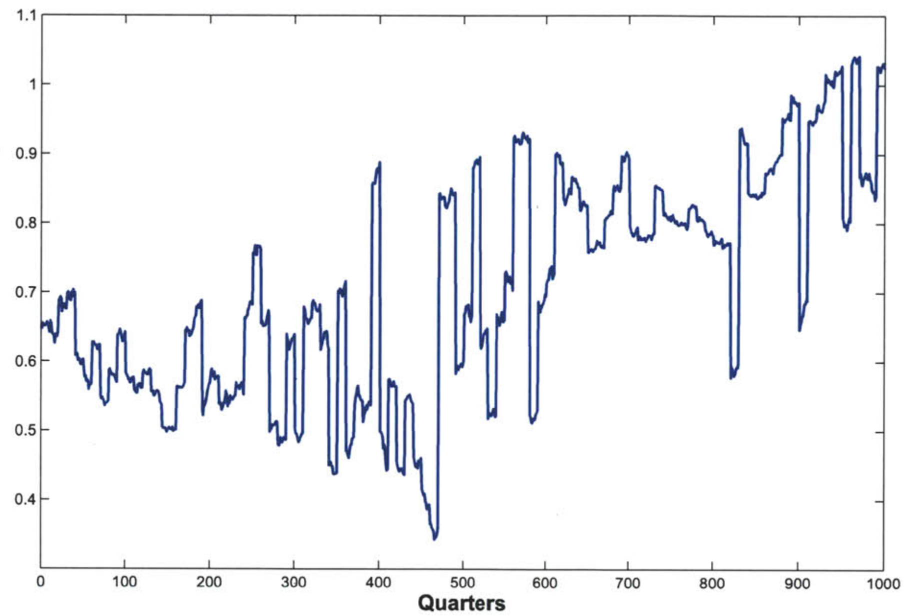
Figure 1: Parameter Estimates



Notes: Variable labels are Psi - ψ, Rho - ρ, Mu g - μ_g, Mu d - μ_d, Sig g - σ_g, Sig d - σ_d, Sig x - σ_x. Horizontal lines indicate the true parameter values. Dashed lines are estimated confidence intervals, averaged pointwise across simulations. The units for the x-axis are quarters. Results are based on 10 simulations of length 1,000.

Finally, taking into account the bias and uncertainty in the agent's beliefs on individual parameters, I consider the speed of convergence for the price of the market asset. Figure 2 displays the ratio of the average across the long simulations of the market price from the model with learning to that under full-information. The market price under learning incorporates all of the agent's uncertainty about both the current state and the true parameter values. The earlier simulations show that this uncertainty induces a greater risk premium, which corresponds to a lower market price. The effect lasts for around 900 quarters or 225 years. Although this effect eventually goes away under the maintained assumptions, if one believes that there are structural changes in the DGP over time, then the effect could last forever, as pointed out by Weitzman (2007).

Figure 2: Market Price Ratio



Notes: The series is computed as the average across simulations of the market price under learning divided by the average market price under the full-information model. Results are based on 10 simulations of length 1,000.

1.6 Empirical

Through simulations, I have shown that introducing uncertainty about parameters into a relatively simple model generates realistic pricing errors and increases the equity premium. The deviations from full information last for more than 200 years, even if the data does not contain structural breaks. In this section, I relax several of my previous assumptions in order to perform a more realistic and direct comparison between the data and the predictions of a model with learning. In particular, I do not impose normality of any shocks nor do I place restrictions on the number of primitive shocks. I also allow agents to use a much richer set of data to uncover the values of any state variables. In this more practical setting, I explore a modification to the moment conditions commonly used to estimate the full information model that allow for learning and discuss how the estimation results change.

1.6.1 Methodology

Recall the agent's Euler equation

$$\begin{aligned} E_t^{\hat{\Pi}_t}\left[u_{j,t+1}\left(\phi_0\right)\right] &= 0 \\ u_{j,t+1}\left(\phi\right) &= \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{j,t+1} - 1 \end{aligned}$$

That is, in equilibrium asset prices must be such that the agent's best forecast of risk-adjusted returns, $u_{j,t+1}$, given all of his information at time t is zero for all assets. This condition ensures that the agent is indifferent between buying and selling any asset, which allows asset markets to clear. Clearly, the moment condition

$$E^{\Pi_0}\left[E_t^{\hat{\Pi}_t}\left[u_{j,t+1}\right]\right] = 0$$

holds. With a model for beliefs, $\hat{\Pi}_t$, one can then use

$$\tilde{u}_{j,t+1} = E_t^{\hat{\Pi}_t}\left[u_{j,t+1}\right]$$

in place of $u_{j,t+1}$ when performing GMM. That is, if we could observe the agent's forecasts of $u_{j,t+1}$, then we could correct the GMM procedure from before. If the preferences and beliefs are correctly specified, the model should have no pricing errors, and the preference parameter estimates should be plausible.

Although I do not observe people's forecasts of an object such as $u_{j,t+1}$, I can use available data to make forecasts based on the most advanced and robust econometric techniques currently available. The standard argument against this approach is that traders have access to a wealth of data that the econometrician does not observe. However, this argument is much less relevant today, as researchers have access to a vastly richer set of aggregate data than nearly 30 years ago when Hansen and Singleton (1982) was written. In fact, the amount of available data has become so large that new statistical methods, such as factor models, have needed to be developed to use all of it accurately. In this section, I assume that the agent is effectively a modern day professional forecaster with access to a vast amount of information about asset markets and the economy as well as modern day tools with which to analyze it and make forecasts.

Generalizing the DGP of the prior section, I assume

$$x_{it} = \mu_i + \lambda_i F_t + e_{it}$$

where $\dim(F_t) = r << N$ is fixed and $X_t = [x_{1t}, \dots, x_{Nt}]$ contains a large set of macroeconomic indicators for the US, including data on aggregate prices, employment, production, housing, and government bond rates. This set of macroeconomic indicators is taken from Sydney Ludvigson's NYU website.[12] I assume further that agents believe

$$u_{j,t+1}(\phi_0) = \alpha_{0,j} + \alpha_{j,1} X_t + \alpha_{j,2} X_t^2 + A u_{j,t}(\phi_0) + B W_t + v_{jt}$$

where e_{it} and v_t are weakly stationary and $T, N \to \infty$. Here, I include the standard three Fama-French factors[13] in W_t. Under these assumptions, Bai and Ng (2007, 2008, 2010) and

[12]The website is located at http://www.econ.nyu.edu/user/ludvigsons.

[13]See Fama and French (1993).

Stock and Watson (2002a, 2002b, 2006) show how to consistently estimate $E_t^{\Pi_0}\left[u_{j,t+1}(\phi_0)\right]$. The procedure is to first estimate the factors, $\hat{F}_t$, by principal components, and then regress $u_{j,t+1}(\phi_0)$ on $\hat{F}_t, \hat{F}_t^2$ as well as lags of $u_{j,t}$ and W_t. This procedure has been widely recommended by the recent literature on forecasting. Stock and Watson (2006) show that such factor based forecasts perform at least as well as other known methods, and Ludvigson and Ng (2008, 2010) have shown that using factors substantially improves predictive power over standard regressors for bond risk premia and the conditional mean and volatility of stock returns. Given its wide ranging support from the forecasting literature, I assume that the agent adopts this method for forecasting $u_{j,t+1}(\phi_0)$.

It is worth comparing this model of learning with the one used for simulations in the previous section. The key difference is that while in the previous section I assumed that the distribution of consumption and dividends is known up to parameter values, I do not make that assumption here. Instead I allow agents to use a more flexible approximation along the lines of a non-parametric regression. One advantage of this approach is that it allows for uncertainty about the functional form of the DGP. Another important practical advantage is that a fully parametric approach is computationally intractible. A thorough search over the parameter space for the relatively simple model of the previous section could easily take two months, even for the exactly identified case.

Although the non-parametric approach has practical advantages, it is worth noting that the model is conceptually somewhat different from before. Previously, by estimating directly the distribution of fundamentals and computing asset prices, the agent was always estimating a correctly specified model. A close examination of the non-parametric model in this section reveals that this is no longer the case. Now, the agent has beliefs directly about asset returns, namely that they are approximately linear in data. As was originally pointed out by Bray and Savin (1986), this amounts to the agent's model being misspecified. The reason is that asset prices are endogenous and depend in particular on the agent's beliefs, which in a model with learning are changing over time. Therefore, when the agent assumes that asset prices follow a particular fixed process, his model must be inherently misspecified as long as he does not have full-information about the distribution of fundamentals. Despite this issue, Bray and Savin (1986) show that the equilibrium still typically converges to the full-information equilibrium. Others have considered such so-called self referential models of learning and found that they have promise for explaining certain features of asset prices,[14] but no work has explored whether such a model can explain pricing errors.

1.6.2 Validation

Although the proposed forecasting method has a strong foundation in the literature, I provide further evidence on its reliability by comparing its performance on key variables with predictions reported in the Survey of Professional Forecasters provided by the Federal Reserve Bank of Philadelphia. The variables that I consider for forecasting are the growth of real personal consumption expenditures[15] and real GDP. These replace $u_{j,t}$ in procedure described above. The forecasting data for the former is available from the third quarter of 1981, and the data for the latter begins on the fourth quarter of 1968. My data on financial and macroeconomic aggregates is only available through the fourth quarter of 2007, so I consider forecasting each series through the first quarter of 2008. I set aside a presample period of 12 quarters to initialize the factor based method. I consider both the BIC and AIC selection criteria for which regressors to include in the forecasting equation, and I also consider performing the regressor selection and parameter estimation in and out of sample.

Table 4 shows the comparison in mean squared error between the forecasts based on the approach that I employ and those made by professionals. One can see that selecting and estimating the forecasting equations out of sample underperforms relative to the professional

[14] See, for example, Adam, Marcet, and Nicolini (2008).

[15] This consumption series is not the same measure used to construct the Euler residuals, as professional forecasts of that series are not publicly available. Forecasting data is available for aggregate consumption while the Euler residuals are computed using consumption of only non-durables and services.

forecasters, suggesting that the selection model has somewhat of a tendency to overfit the data. If both the selection and estimation are done in sample, the factor based procedure is, not surprisingly, superior to professional forecasters. Meanwhile if the selection is done in sample while the estimation is done out of sample, the model does roughly as well as professional forecasters. Using the AIC criteria amplifies the problem of overfitting when compared with BIC. For these reasons, I report my main results in terms of the BIC, and consider the procedure with selection in sample and estimation out of sample as my baseline. In the appendix, I show further that the employed factor based forecasting method outperforms simpler autoregressions.

Table 4: Forecast Comparison

	MSE		**In / Out of Sample**	
	GDP	**Cons. Growth**	**Selection**	**Estimation**
Pro. Avg.	6.2131e-4	3.1342e-4	-	-
BIC	7.5113e-4	4.6494e-4	Out	Out
BIC	6.5591e-4	2.6245e-4	In	Out
BIC	5.782e-4	1.8532e-4	In	In
AIC	7.8148e-4	4.6501e-4	Out	Out
AIC	8.7754e-4	2.8125e-4	In	Out
AIC	4.6722e-4	1.263e-4	In	In

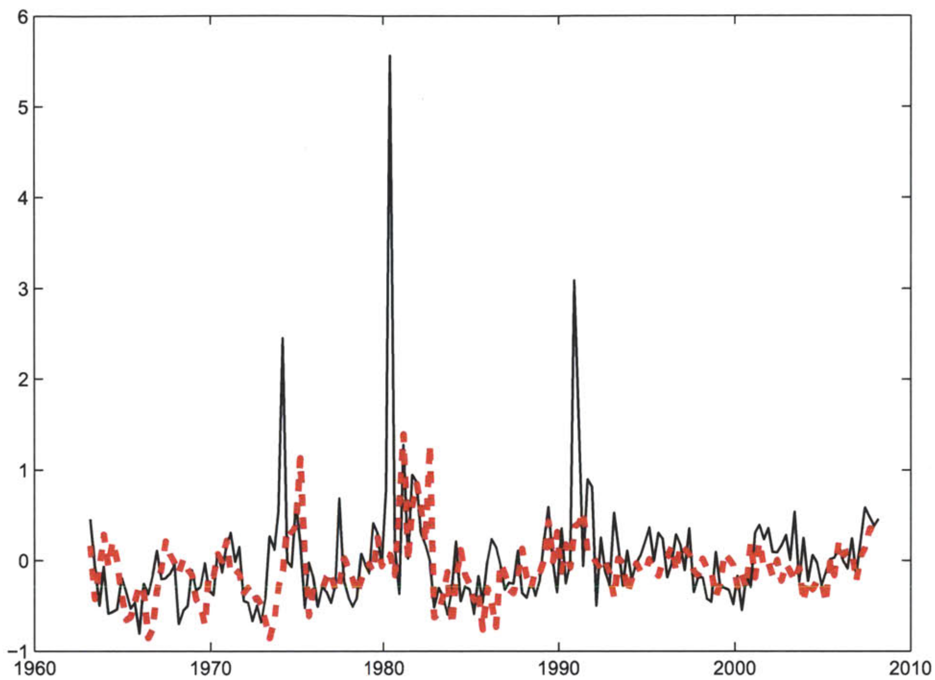
1.6.3 Estimation

Finally, I estimate the model by GMM, replacing the typical moment condition that uses ex-post returns with the analagous condition from my model that involves ex-ante forecasts. Based on the findings of my prior section, my benchmark analysis uses BIC to select the regressors to include in the prediction model using the entire sample and estimate the coefficients in the regression out of sample. I start by presenting the evidence on pricing errors; that is, for GMM with an identity weighting matrix. I consider two sets of assets one including only the market and risk-free returns and one including also the 6 Fama-French size and book to market sorted portfolios. I omit the period 1960-1962, the first 12 quarters of my data, from the estimation to serve as a pre-period for initializing the forecasts.

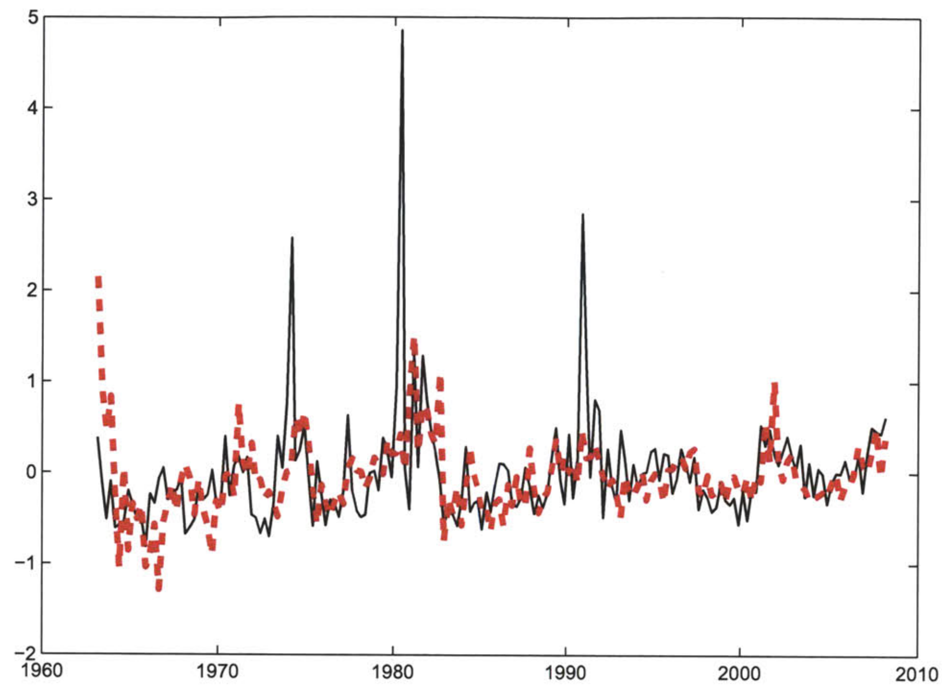
Before getting to the estimation, Figure 3 compares the realized and forecasted pricing errors when both are computed at the full information estimates for the exactly identified case, reported in Table 1. For the forecasted pricing errors, I use the factor based procedure with selection in sample and estimation out of sample. One can see that, consistent with the findings of LL, several outliers corresponding to troughs of US recessions explain a substantial fraction of the full information model's pricing errors, though removing these dates does not reverse the statistical rejection of the model in the overidentified case. Interestingly, the model with learning avoids those outliers for the most part, as such rare spikes are essentially unforecastable.

Figure 3: Pricing Errors

Panel A: R_m



Panel B: R_f



Notes: Parameters are set at $\gamma = 111$ and $\beta = 1.59$, the full information estimates for the exactly identified case. The solid line is for the full information case and the dashed line is for the model with learning where selection is in sample and estimation is out of sample.

Table 5 presents the results on the formal estimation of the model with learning for the identity weighted GMM objective. Panel A shows the estimates for the exactly identified system with only the risk-free rate and the market return. The first line is the same as the first line of Table 1, and shows the estimates for the full information model. The second through fourth rows show the estimates for the model with learning under three different scenarios for whether selection and estimation are performed in or out of sample. As discussed previously, the full information model leads to implausibly high estimates for risk aversion and large pricing errors. When using forecasts of Euler residuals instead of realized values, the pricing errors disappear completely. This finding is robust to selecting the regressors in or out of sample and similarly for the estimation. Except in the case of performing both selection and estimation in sample, the estimates of both the risk aversion and time discount factors also become much smaller and realistic, though they do not fall by enough to make them economically acceptable. It is not very surprising that the model with selection and estimation done in sample is more similar to the full information case than the others. It uses more information than a person would realistically have, and as shown above, it overfits the data.

In Panel B of Table 5, I show the results for the overidentified system with 8 test assets that are chosen to be representative of the cross section of US equities. The pricing errors of the model decline when using forecasts. While the full information model is rejected at the 1% significance level by an overidentification test, none of the versions of the model with learning can be rejected.[16] For the first two models of learning, the pricing errors decline by roughly one half, and the estimated risk aversion declines by more than one half. These results are consistent with those in Panel A. Broadly, the findings in Table 5 are consistent with the earlier simulation results, and suggest that learning can account for the full information model's pricing errors and a portion of the equity premium puzzle.

[16]The second stage, efficient GMM, estimates that were used to compute the overidentification test statistic are reported in the appendix in Table A4.

Table 5: Estimates

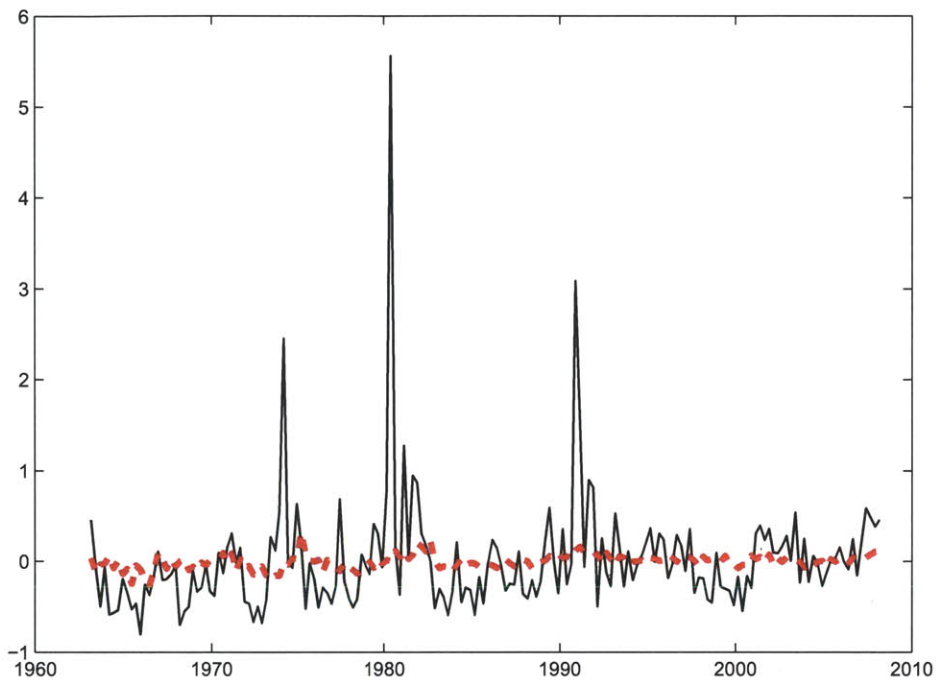
Model	$\hat{\gamma}$	$\hat{\beta}$	RMSE	$\frac{RMSE}{RMSR}$	Selection	Estimation
Panel A: Exactly Identified						
Full Information	111	1.59	2.12%	0.41	-	-
Learning	19	1.10	0.00%	0	Out	Out
Learning	25	1.14	0.00%	0	In	Out
Learning	118	1.61	0.00%	0	In	In
Panel B: Overidentified						
Full Information	92	1.48	3.33%	0.35**	-	-
Learning	25	1.14	1.50%	0.16	Out	Out
Learning	45	1.28	1.57%	0.17	In	Out
Learning	142	1.74	2.98%	0.32	In	In

Notes: Results based on an equally weighted objective (i.e., $W = I$). In the overidentified case, stars indicate whether the model is rejected (* - rejected at the 5% level, ** - rejected at the 1% level). The test of overidentification uses a Newey-West covariance matrix.

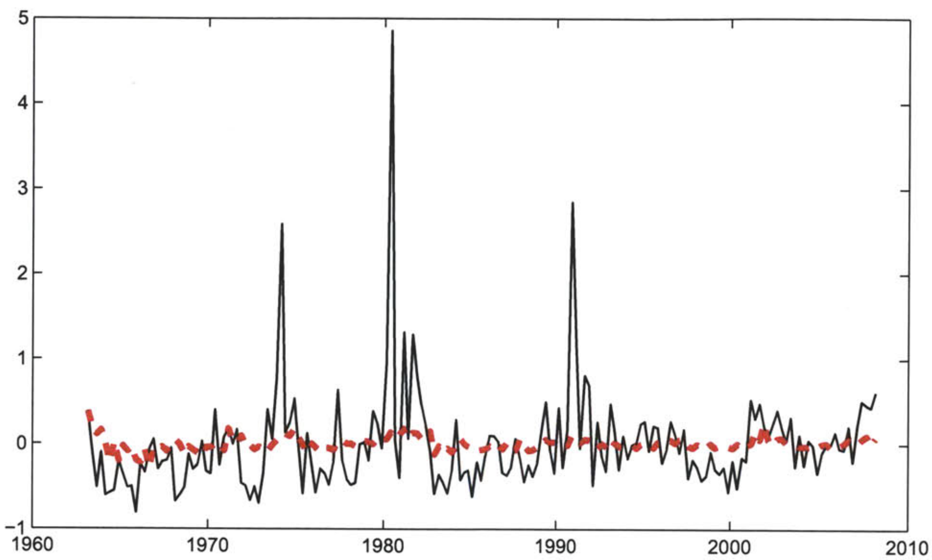
In Figure 4, I plot the estimated Euler residuals for the market return and risk-free rate for the models with learning and full information. The plotted series correspond to the respective optima of the two models on the unrestricted parameter space; that is, they are not computed for the same parameter values, unlike Figure 3. Forecasts are shown for the baseline case of performing model selection in sample and estimation out of sample. Figure 4 reveals that the optimized forecasted residuals are much smaller and much less volatile than the optimized realized residuals. Therefore, the model with learning cannot be rejected, despite the fact that its moments are much less volatile than those of the full information model.

Figure 4: Pricing Errors

Panel A: R_m



Panel B: R_f



Notes: Parameters are set at $\gamma = 111$ and $\beta = 1.59$, the full information estimates for the exactly identified case. The solid line is for the full information case and the dashed line is for the model with learning where selection is in sample and estimation is out of sample. For the full information series, $\gamma = 111$ and $\beta = 1.59$. For learning, $\gamma = 25$ and $\beta = 1.14$.

1.7 Conclusion

The recent work of LL proves that existing full information asset pricing models cannot explain why the standard consumption-based model with CRRA preferences is rejected for large values of risk aversion. In this paper, I consider learning as an explanation. I provide a simple example which illustrates that LL's main theorem does not extend to the case when agent's have uncertainty about underlying DGP parameters. I then construct a model of learning in a more general environment and provide simulation evidence that it can generate quantitatively realistic pricing errors and a substantial equity risk premium, in contrast to the full information version of the model. Finally, I estimate a model with learning and find that it is not rejected in the data, producing pricing errors that are statistically indistinguishable from zero.

1.8 Appendix

1.8.1 Data Summary Statistics

Table A1: Summary Statistics

	Δc_t	Δd_t	R_m	R_f	R_1	R_2	R_3	R_4	R_5	R_6
Mean (%)	0.51	0.30	1.78	0.34	1.85	3.09	3.63	1.67	1.90	2.51
Std. Dev. (%)	0.43	11.04	8.32	0.57	13.70	10.75	11.01	8.96	7.55	8.06

Notes: Δc_t - log consumption growth; Δd_t - log dividend growth; R_m - market return; R_f - risk-free return; R_1,...,R_6 - returns to Fama French portfolios sorted on sized and book to market.

1.8.2 Additional Simulations

Table A2: Simulation Results

Object	$E[r_m - r_f]$	$\sigma[r_m]$	$E[r_f]$	$\sigma[r_f]$	$Q\left(\frac{RMSE}{RMSR}, 90\%\right)$	$Q\left(\frac{RMSE}{RMSR}, 95\%\right)$
Model	1.76%	20.70%	2.90%	0.93%	0.14	0.22
Full Info.	0.04%	11.60%	2.89%	1.18%	0.05	0.07
Data	1.44%	8.32%	0.34%	0.57%	0.41	0.41

Notes: Preference parameters are $\beta = 0.997$, $\gamma = 5$. DGP parameters are set at the maximum likelihood estimates from Table 2. $Q(X, y)$ denotes the y-th quantile of X.

1.8.3 Forecast Comparisons

Here, I consider how well the factor based procedure works at forecasting the Euler residuals both in and out of sample as compared with simpler autoregressions. Table A3 reports the ratio of the mean squared prediction error to the variance of the realized values. Selection of regressors in the factor based approach is done in sample using BIC in all cases. Naturally, in sample forecasts are much better than for a simpler AR model, as the proposed methodology has more degrees of freedom. More significantly, the proposed procedure tends to outperform the simpler AR model out of sample as well. For small values of γ, the Euler residuals become nearly unforecastable under either approach. This is not surprising as Goyal and Welch (2008) find that returns are unforecastable out of sample. On the other hand, for larger values of γ, the Euler residuals become much more forecastable, and the proposed methodology dominates the simpler AR model. This provides another piece of evidence, on top of an already large literature, that factor based forecasts perform at least as well as other commonly used models.

Table A3: Forecast Performance

γ	**5**	**10**	**25**	**50**	**75**	**100**	**125**	**150**
			Panel A: Factor-Based (Out of Sample)					
R_m	1.0279	0.9648	0.9112	0.8919	0.8956	0.9210	0.9369	0.9530
R_f	0.9390	0.9446	0.9425	0.9684	0.9780	0.9928	1.0008	1.0021
R_1	1.0607	1.0481	0.9099	0.8660	0.8812	0.8986	0.9241	0.9562
R_2	1.0574	1.0283	0.9022	0.8981	0.8988	0.9338	0.9346	0.9719
R_3	1.0904	1.0574	0.9090	0.9314	0.9052	0.9195	0.9386	0.9544
R_4	1.0045	1.0181	0.8836	0.8684	0.8983	0.9289	0.9494	0.9451
R_5	1.0025	0.9495	0.9347	0.9001	0.9237	0.9375	0.9443	0.9599
R_6	1.0257	0.9596	0.9413	0.9203	0.9022	0.9291	0.9442	0.9601

Panel B: AR(6) (Out of Sample)								
R_m	1.1423	1.1116	1.3147	1.0259	1.0571	1.0406	1.1745	1.2261
R_f	0.9077	1.0195	0.9842	0.9824	1.0128	1.1200	1.1107	1.2304
R_1	1.0605	1.0696	1.1371	1.0657	1.1318	1.0637	1.1421	1.2405
R_2	1.1297	1.1122	1.0964	1.0167	1.0169	1.0529	1.1592	1.2603
R_3	1.1128	1.1298	1.2446	1.0427	1.0411	1.0797	1.1830	1.2400
R_4	1.2043	1.0146	1.0355	1.0180	1.0493	1.0159	1.1575	1.2023
R_5	1.1849	1.2209	1.0337	1.2297	1.0636	1.0541	1.1590	1.2623
R_6	1.1374	1.1548	1.0795	1.0542	0.9997	1.0477	1.1666	1.2570
Panel C: Factor-Based (In Sample)								
R_m	0.9515	0.9233	0.7862	0.7740	0.7812	0.7988	0.8203	0.8428
R_f	0.7902	0.8034	0.8042	0.8126	0.8259	0.8430	0.8622	0.8817
R_1	0.9447	0.9241	0.7865	0.7358	0.7658	0.7870	0.8126	0.8381
R_2	0.9418	0.9157	0.7788	0.7412	0.7725	0.7903	0.8115	0.8333
R_3	0.9285	0.9231	0.7996	0.7560	0.7827	0.7972	0.8161	0.8361
R_4	0.9502	0.9237	0.8191	0.7651	0.7736	0.7925	0.8152	0.8385
R_5	0.9595	0.9317	0.8016	0.7856	0.7888	0.8037	0.8235	0.8448
R_6	0.9478	0.9175	0.8034	0.7691	0.7947	0.8079	0.8264	0.8466

Panel D: AR(6) (In Sample)								
R_m	0.9945	0.9945	0.9024	0.8139	0.8044	0.8240	0.8560	0.8909
R_f	0.7474	0.7523	0.7350	0.7510	0.7667	0.7945	0.8307	0.8696
R_1	0.9945	0.9946	0.9946	0.8864	0.8456	0.8488	0.8721	0.9017
R_2	0.9946	0.9946	0.9946	0.8566	0.8292	0.8395	0.8661	0.8972
R_3	0.9945	0.9946	0.9946	0.8714	0.8410	0.8479	0.8716	0.9005
R_4	0.9945	0.9945	0.8981	0.8091	0.7988	0.8184	0.8510	0.8868
R_5	0.9945	0.9945	0.8912	0.8099	0.8035	0.8240	0.8563	0.8913
R_6	0.9945	0.9945	0.9568	0.8376	0.8213	0.8350	0.8624	0.8940

Notes: Reported statistics are the mean squared forecasting error divided by the variance of the underlying series. I set $\beta = 0.997$.

1.8.4 Efficient Estimation

Table A4: Optimally Weighted Estimates

Model	$\hat{\gamma}$	$\hat{\beta}$	Selection	Estimation
Panel A: Exactly Identified				
Full Information	111	1.61	-	-
Learning	17	1.09	Out	Out
Learning	25	1.14	In	Out
Learning	118	1.61	In	In
Panel B: Overidentified				
Full Information	83	1.47	-	-
Learning	16	1.08	Out	Out
Learning	25	1.14	In	Out
Learning	173	1.85	In	In

Notes: Results based on an optimally weighted objective (i.e., $W = \hat{\Sigma}^{-1}$). In the overidentified case, stars indicate whether the model is rejected (* - rejected at the 5% level, ** - rejected at the 1% level). The test of overidentification uses an optimal weighting matrix that is robust to heteroskedasticity and autocorrelation.

1.8.5 Proofs

Proof of Theorem 1 By direct computation,

$$
\begin{aligned}
E^{\Pi_0}\left[\beta\exp\left(-\gamma\Delta c_{t+1}\right)R_{t+1}\right] &= 1 \\
&\Leftrightarrow \\
E^{\Pi_0}\left[\beta\exp\left(-\gamma\Delta c_{t+1}\right)\begin{pmatrix}\frac{1}{P_{f,t}}\\ \frac{D_{t+1}}{P_{1,t}}\end{pmatrix}\right] &= 1 \\
&\Leftrightarrow \\
E^{\Pi_0}\left[\beta\exp\left(-\gamma\Delta c_{t+1}\right)\begin{pmatrix}\frac{1}{\beta_0 E_t^{\hat{\Pi}_t}\left[\exp(-\gamma_0\Delta c_{t+1})\right]}\\ \frac{D_{t+1}}{\beta_0 E_t^{\hat{\Pi}_t}\left[D_{t+1}\exp(-\gamma_0\Delta c_{t+1})\right]}\end{pmatrix}\right] &= 1
\end{aligned}
$$

$$\Leftrightarrow$$

$$
E^{\Pi_0}\left[\frac{1}{\beta_0 E_t^{\hat{\Pi}_t}\left[\exp\left(-\gamma_0\Delta c_{t+1}\right)\right]}\right] = E^{\Pi_0}\left[\frac{D_{t+1}}{\beta_0 E_t^{\hat{\Pi}_t}\left[D_{t+1}\exp\left(-\gamma_0\Delta c_{t+1}\right)\right]}\right]
$$

$$
\begin{aligned}
&\Leftrightarrow \\
E^{\Pi_0}\left[\frac{D_{t+1}}{E_t^{\hat{\Pi}_t}\left[D_{t+1}\right]}\right] &= 1 \\
&\Leftrightarrow \\
\mu_{d,0}\left(1+\sigma^2 t\right)E^{\Pi_0}\left[\left(z_d+\sigma^2\sum_{j=1}^{t}D_j\right)^{-1}\right] &= 1
\end{aligned}
$$

where I utilize the fact that C_t and D_t are iid.

Chapter 2

High-Dimensional Impulse Response Function Matching Estimators[1]

2.1 Introduction

Impulse response function matching is a widely used method for estimating the structural parameters of dynamic stochastic general equilibrium (DSGE) models.[2] The *common* impulse response function matching estimator (IRFME) computes estimates of structural parameters by minimizing the distance between the IRFs estimated from the data via structural vector autoregressions (SVARs) and the corresponding true coefficients implied by the model for given parameters. As a limited information approach, it has the advantage that it does not require a full and correct specification of the data generating process. It is also relatively simple to implement compared to maximum likelihood. Despite its widespread usage, the properties of the common IRFME have not been established for typical settings where there are a large number of impulse response coefficients being matched.

[1]This paper is co-authored with Anna Mikusheva.

[2]The method was first used by Rotemberg and Woodford (1997), and more recent examples include Altig et al (2005, ACEL hereafter), Christiano et al (2005), DiCecio (2005), Iacoviello (2005), Boivin and Giannoni (2006), Uribe and Yue (2006), DiCecio and Nelson (2007), and DuPor, Han and Tsai (2009).

This paper is the first to consider the properties of the common IRFME when the data generating process is an infinite order VAR and the horizon of impulses being matched is non-negligible relative to the sample size. Both of these are typical features encountered by those estimating DSGE models. For example, Altig et al (2005, hereafter ACEL) uses 170 quarters of data and matches impulse responses up to a horizon of 20 quarters when their theoretical specification is VARMA (i.e., infinite order VAR).[3][4] We begin by showing that the common IRFME is still consistent and asymptotically normal in this setting, as long as two important conditions are satisfied. First, the order of the estimated VAR must grow with the sample size but slowly enough to allow for consistent estimation of IRF coefficients. Second, the number of IRFs being matched must be of the same order as the number of lags in the estimated VAR. However, this second condition is rarely close to true in practice. For example, ACEL estimates a VAR with 4 lags and matches IRFs up to a horizon of 20.[5]

Next, we construct a simple example where the horizon of matched IRFs increases too rapidly, and the common IRFME is not consistent. We show that consistency breaks down because of bias in the estimation of IRF coefficients at long horizons. In general, bias in standard IRF coefficient estimates comes from three potential sources: 1) the non-linearity of the mapping between VAR and IRF coefficients; 2) finite sample bias in VAR coefficient estimates; and 3) the fact that a finite order VAR is estimated whereas the true model has infinite lags. Prior work has suggested that the third source of bias may be substantial. In our example, we go further by showing that even when the estimated model is correctly

[3] Altig et al (2011) is a more recent version of this paper, which uses a sample with only 110 quarters of data. Since this paper was initially written prior to the release of the updated version of ACEL, we refer to the setting and parameter estimates from the earlier version throughout.

[4] Christiano et al (2005) uses 124 quarters of data and matches IRFs up to a horizon of 25. DiCecio (2005) uses 170 quarters and matches IRFs at a horizon of up to 20. Iacoviello (2005) uses 118 quarters of data and matches IRFs up to a horizon of 20. Boivin and Giannoni (2006) use two samples, one with 82 quarters and one with 92 quarters, and matches IRFs up to a horizon of 16. Uribe and Yue (2006) use 32 quarters of data and matches IRFs up to a horizon of 24. DiCecio and Nelson (2007) use 107 quarters of data and matches IRFs up to a horizon of 25. DuPor, Han, and Tsai (2009) use 195 quarters of data and matches IRFs up to a horizon of 20. None of these papers estimate a VAR with more than 4 lags in their baseline specification.

[5] See footnote 2 for more examples.

specified, the first source of bias alone is enough to generate inconsistency. We show that at longer horizons, the first type of bias in IRF coefficient estimates becomes large relative to typical estimates of those coefficients' variances, the inverses of which are used as weights in the objective function. These two confounding factors eliminate the consistency of the common IRFME.

Finally, we examine the properties of the common IRFME in the realistic setting of ACEL, and we compare them with those of an alternative, less commonly used "bias corrected" IRFME. The bias corrected approach is based directly on the method of indirect inferences, as in Gourieroux et al (1993). It compares the IRFs estimated from the data with the average of those estimated in the same way from model simulated data for given parameters. Intuitively, this controls for any potential bias in the finite sample IRF estimates. In the setting considered by ACEL, we find that the common IRFME has large bias and is far from normally distributed. In contrast, the bias corrected estimator has lower bias, and it produces parameter estimates that reverse ACEL's main conclusion. Specifically, ACEL finds that firm-specific capital is key for reconciling aggregate inflation dynamics and firm-level evidence on the frequency of price adjustment. However, estimates from the bias corrected approach imply that a model with firm-specific capital does no better than a model with homogeneous capital in this regard. Moreover, we find that the reported confidence intervals for the key parameter, which assume its estimate is normally distributed, are much smaller than those obtained from the parametric bootstrap, which include the entire relevant parameter space. Therefore, our findings suggest that the common IRFME and the associated commonly reported standard errors have poor properties in the practical settings where they have recently been applied. We find that the bias corrected approach, based on indirect inferences, provides an improvement.

This paper is related to two main strands of literature. First, it is related to the recent work on the properties of the common IRFME. Dridi et al (2007) discusses the properties of general indirect inference estimators for DSGE models. We extend their results to the case where the number of IRFs being matched grows unboundedly with the sample size. In a similar setting to ours, Jorda and Kozicki (2007) derive the properties of an alternative IRFME, based on the local projections approach of Jorda (2005). However, Kilian and Kim (2009) argue that the approach of Jorda (2005) actually produces an estimator with higher bias and variance than standard approaches. Hall et al (2010) proposes a selection criteria for choosing the number of IRFs to match, but assume that this is bounded as the sample size grows.

Second, we contribute to a recent debate on the accuracy of the common IRFME in pratical settings. This concern was recently raised by Chari et al (2008), which cited the fact that the typical approach employs a VAR with a small number of lags, while the model implies an infinite order VAR. Those authors argue that the problem is especially bad when SVARs are estimated using long-run restrictions, as in ACEL, and they recommend using the bias corrected estimator on intuitive grounds, referring to it as the Sims-Cogley-Nason approach.[6] In response, Christiano et al (2007) show that in an alternate specification to the one considered by Chari et al (2008), SVARs have good finite sample properties, and argue that the cases where SVARs perform poorly are rejected by the data. We contribute to this debate in two ways. First, existing arguments have centered around the misspecification of the estimated VAR, the third source of bias listed above. We show that even in the absence of misspecification, the common IRFME can still be inconsistent in the realistic case where many IRF coefficients are matched. Second, we show that in a recent practical example, the common IRFME is badly biased, and the bias corrected approach provides an improvement.

The rest of the paper will be structured as follows: Section II explains the basics of the two approaches to IRF matching that we consider; Section III extends existing results on the consistency and asymptotic normality of the common approach to settings where the

[6]The title "Sims-Cogley-Nason" was inspired by the work of Sims (1989) and Cogley and Nason (1995).

data generating process is VARMA and the number of IRFs being matched grows with the sample size; Section IV describes a simple example where the number of IRFs being matched grows too quickly and common approach is inconsistent; Section V compares the common and bias corrected IRFMEs for the ACEL model.

2.2 Impulse Response Function Matching Estimators

One way of estimating a DSGE model is by matching the IRFs produced by the model with those estimated from the data. In recent years, IRF matching has gained popularity because it is relatively simple to implement and flexible enough to accomodate many models. As a form of calibration, this estimator is used when the econometrician believes that their model is too simplistic to produce a valid likelihood but rich enough to correctly describe some dynamic features of the data. We begin by outlining the general framework for estimating DSGE models via calibration and treat IRF matching as a special case.

Before estimation can be performed, one must rewrite the model into a workable form. The pre-econometric stage includes "approximately solving the model" and usually writing it in terms of a state-space representation. As the output of this stage, one typically has the following system:

(a) Measurement Equation

$$X_t = G(X_{t-1}, s_t, \beta);$$

(b) State Equation

$$s_t = g(s_{t-1}, X_{t-1}, u_t, \beta).$$

Here s_t is a state variable, X_t is a vector of all observed variables, u_t are shocks, and β is a vector containing the structural parameters of interest.

With the state-space representation in hand, one must specify the features of the data that the model explains (e.g., IRFs, covariance structure). A comparison between model predictions and estimates of the specified features of the data serves as the criterion for estimating β. Assuming that one can simulate $X_T(\beta)$, a sample of size T given the parameter vector β, from the model, the structural parameters can be estimated directly by Simulated GMM (Indirect Inferences). In doing so, it is acknowledged that the model may not produce a completely valid likelihood due, for example, to a misspecified distribution of the error term because otherwise one would use a likelihood based approach.

We denote $Y_T = \{Y_{i,t}, i = 1, ..., n; t = 1, ..., T\}$ to be the data where $Y_T \sim F_T$. We do not assume that F_T is the same as the distribution $G_T(\beta)$ for the simulated data $X_T(\beta)$, but some characteristics (e.g., IRFs), denoted by $\theta\left(.\right)$, of the two distributions, F_∞ and $G_\infty\left(\beta\right)$, are the same at the true value of the structural parameter, β_0. Fixing notation, we define $\theta_0 \equiv \theta(F_\infty)$ and assume that $\theta_0 = \theta(G_\infty(\beta_0)) = \theta(\beta_0)$. Typically, one can calculate $\theta(\beta)$ either by running long simulations or directly from the state-space representation, (a) and (b).

2.2.1 The Common Approach

Given the above environment, estimation of β typically is as follows:

- **First stage:** Using the data, Y_T, estimate the reduced form parameter (feature) as $\widehat{\theta}\left(Y_T\right)$ such that $\widehat{\theta}\left(Y_T\right) \rightarrow^p \theta_0$ as $T \rightarrow \infty$. If one wishes to construct confidence sets for β, the first stage estimator, $\widehat{\theta}$, must also be asymptotically normal.

- **Second stage:** Compute the estimator $\hat{\beta}$ by minimizing a measure of the distance between the estimated reduced form parameter, $\widehat{\theta}\left(Y_T\right)$, and the corresponding true parameter implied by the model, $\theta(\beta)$. The distance metric is usually quadratic, and one may use the "GMM optimal weighting matrix." This produces

$$\hat{\beta} = \arg\min_{\beta} \left(\hat{\theta}\left(Y_T\right) - \theta\left(\beta\right)\right)' W \left(\hat{\theta}\left(Y_T\right) - \theta\left(\beta\right)\right).$$

When $k = dim(\theta_0)$ is finite, the consistency of $\hat{\beta}$ follows from a two-stage logic: if the first stage is consistent and asymptotically normal, then the second stage will also be consistent and asymptotically normal. This two-stage argument is analogous to "minimal distance asymptotics."

The common approach to IRF matching is exactly this procedure where $\hat{\theta}(Y_T)$ refers to IRFs estimated from the data, and $\theta(\beta)$ refers to the corresponding IRFs computed directly from the state-space representation of the model for a given value of β. Under the stated assumptions and as long as k is finite, we agree that this method is asymptotically valid; however, in finite samples when k is large relative to T, $\hat{\theta}(Y_T)$ may be badly biased which in turn may cause $\hat{\beta}$ to be badly biased. Similarly, when $k \to \infty$ as $T \to \infty$, this bias may not go away asymptotically either. In short, although there is a valid argument that $\hat{\beta} \to^p \beta_0$ as $T \to \infty$ for k fixed, it does not imply that $\hat{\beta} \to^p \beta_0$ as $\frac{T}{k} \to c > 0$ or $E\left(\hat{\beta}(Y_T)\right) \approx \beta_0$ for k large relative to T. In practice, it is often the case that k is large relative to T, as described in the introduction. Furthermore, DSGE models theoretically match the IRFs at all horizons, so for efficiency, one would reasonably like to have $k \to \infty$ as $T \to \infty$.

2.2.2 The Bias Corrected Approach

The modified procedure we propose is

- **First stage:** This is the same as before. Compute $\widehat{\theta}(Y_T)$.
- **Second stage:** Instead of computing $\theta(\beta)$ as before, proceed as follows:
  - Simulate many data sets $\{X_T^i(\beta)\}_{i=1}^M$ from the model, (a) and (b), where M is the number of simulated samples. The sample size for each simulated data set is T, the same as in the actual data.
  - Compute $\widehat{\theta}(X_T^i(\beta))$ using exactly the same estimator as in the first stage, and

finally compute

$$\hat{E}\left(\widehat{\theta}\left(X_T\left(\beta\right)\right)\right) = \frac{1}{M}\sum_{i=1}^{M}\widehat{\theta}\left(X_T^i\left(\beta\right)\right).$$

- Define the estimator, $\tilde{\beta}$, as

$$\tilde{\beta} = \arg\min_{\beta}\left(\hat{\theta}\left(Y_T\right) - \hat{E}\left(\widehat{\theta}\left(X_T\left(\beta\right)\right)\right)\right)' W\left(\hat{\theta}\left(Y_T\right) - \hat{E}\left(\widehat{\theta}\left(X_T\left(\beta\right)\right)\right)\right)$$

Intuitively, as long as $\widehat{\theta}\left(X_T\left(\beta_0\right)\right) \rightarrow^p \theta_0$ as $T \rightarrow \infty$ and the assumptions for the consistency of the common approach hold, the bias corrected approach is also consistent. On the other hand, in some settings, it may have more desirable properties, such as lower bias. For example, consider an estimator $\hat{\theta}$ such that $E\left(\hat{\theta}\left(Y_T\right)\right) = B\frac{k}{T} + \theta_0$ and $E\left(\widehat{\theta}\left(X_T\left(\beta_0\right)\right)\right) = B\frac{k}{T} + \theta_0$; then on average the bias from first stage estimation cancels out in the objective function leading to an unbiased estimator $\tilde{\beta}$. In other words, the idea is to compute the IRFs from the model in the same way that we compute them from the data. That way, if the first stage is biased, then the second stage will have a similar bias that will cancel out in the objective function.

2.3 Consistency with Many Reduced Form Parameters

In this section, we extend previous work on the asymptotic properties of the common approach by considering the realistic case where the data does not have a finite order VAR representation and the number of IRF coefficients being matched is not negligible relative to the sample size.[7] In particular, a typical solution to a DSGE model has VARMA form and usually cannot be written as an VAR of finite order. Recent empirical papers using the common IRFME also usually match a large number of IRFs, such as ACEL which matches

[7]Existing proofs for the consistency and asymptotic normality of IRFMEs are contained in Dridi et al (2007) and Hall et al (2010). These papers assume that the data generating process has a VAR(p) representation for a finite p, and the number of IRFs being matched is finite.

IRFs at horizons of up to 20 periods.[8] As a result, the setting we consider provides an improved approximation to situations commonly encountered in the estimation of DSGE models.

2.3.1 Setup

Assume that we have data on the process $\{y_t\}_{t=1}^{T}$, where y_t is K-dimensional and has a VAR(∞) representation

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t.$$

Here ε_t are i.i.d. with $E(\varepsilon_t) = 0.$. Assume that the process y_t also has the following MA(∞) representation $y_t = \sum_{j=0}^{\infty} \Theta_j \varepsilon_{t-j}$. Here Θ_j is a set of impulse responses of K variables, y_t, to K shocks at horizon j, ε_{t-j}. Assume we have a theoretical model that produces theoretical impulse responses, $\Theta_j(\beta)$, for any given value of a vector of structural parameters, β, and for any horizon j. Assume that the true value of the coefficients is β_0 and $\Theta_j(\beta_0) = \Theta_j$. We are concerned with estimating β_0 through matching IRFs from the data with those from the model.

In practice, estimation of this model is typically done by estimating a finite order VAR(r), $y_t = \sum_{j=1}^{r} A_{j,r} y_{t-j} + e_t$ and assuming that the order of VAR $r = r_T$ is increasing with sample size. This approach is commonly referred to as sieve-VAR. We denote the OLS coefficient estimates of the VAR(r) as $\hat{A}_{j,r}$. Then impulse responses are estimated as

$$\widehat{\Theta}_{j,r} = \sum_{i=1}^{j-1} \widehat{\Theta}_{i,r} \widehat{A}_{j-i,r},$$

where $\hat{A}_{j,r} \equiv 0$ for all $j < 0$.

Finally, the structural parameter vector,β_0, is estimated by matching estimated IRFs with those from the model. For this section, we assume that one matches all impulse responses

[8] See footnote 2 for more examples.

at horizons up to q. Let $\hat{\theta}_{q,r} \equiv vec\left[\widehat{\Theta}_{1,r} : \ldots : \widehat{\Theta}_{q,r}\right]$ be the $(K^2q) \times 1$ vector of estimated IRF coefficients to be matched and $\theta_q(\beta) \equiv vec\left[\Theta_1(\beta) : \ldots : \Theta_q(\beta)\right]$ be the corresponding theoretical IRF coefficients from the model for a given β. Estimation is done by minimizing a quadratic function of the difference between the model produced impulse responses and the ones estimated from data, as discussed in the preceding section. Let Q be a symmetric semi-positive definite $(K^2q) \times (K^2q)$ matrix. The objective function is

$$\lambda(\beta) = (\widehat{\theta}_{q,r} - \theta_q(\beta))'Q(\widehat{\theta}_{q,r} - \theta_q(\beta)).$$

The estimator is

$$\widehat{\beta} = argmin_{\beta}\lambda(\beta).$$

We make the following two assumptions.

Assumption 1

i Let the coefficients for the VAR(∞) and MA(∞) representations of the process y_t satisfy the following conditions: $\left|\sum_{j=1}^{\infty} \Theta_j z^j\right| \neq 0$ for $|z| \leq 1$, $\sum_{j=1}^{\infty} \|A_j\| < \infty$, and $\sum_{j=1}^{\infty} \|\Theta_j\| < \infty$ where $\|A\| = \sqrt{tr(A'A)}$.

ii ε_t are independent and identically distributed with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$,

$E(|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}|) < \infty$, and $1 \leq i, j, k, l \leq K$.

iii $r = r_T$ where $r_T \to \infty$ and $r_T^3/T \to 0$ as $T \to \infty$.

iv $\sqrt{T}\sum_{j=r_T+1}^{\infty} \|A_j\| \to 0$ as $T \to \infty$.

Assumption 2

i $\lim_{q\to\infty}(\theta_q(\beta_0) - \theta_q(\beta))'Q(\theta_q(\beta_0) - \theta_q(\beta)) = f(\beta) < C < \infty$ uniformly over β.

ii For any $\varepsilon > 0$,

$$\inf_{|\beta-\beta_0|>\varepsilon} f(\beta) > f(\beta_0) = 0.$$

iii $q = q_T = r_T$.

iv For any β, $\sum_{j=1}^{\infty} \Theta_j(\beta) < \infty$.

Assumption 1 is taken from past work on sieve-VAR.[9] The former shows that Assumption 1 is sufficient to ensure the consistency and asymptotic normality of any finite dimensional vector of IRF coefficient estimates. The latter shows that under these conditions, any linear combination of IRF coefficient estimates is consistent and asymptotically normal.[10] Importantly, condition (iii) requires that the number of lags of the estimated AR process does not increase too quickly with the sample size. This is in line with ACEL, our primary empirical example, where $r = 4$ and $T = 170$ in the benchmark estimation.

Assumption 2 provides the additional conditions we require for the consistency of the IRF matching estimator of β. Conditions (i) and (ii) are standard and guarantee that the objective function is bounded over the parameter space and unique minimized at the true parameter value. Similarly, condition (iv) requires that the IRF coefficients be summable across the entire parameter space. Most importantly, condition (iii) requires that the number of matched IRF coefficients grows no faster than the number of estimated AR coefficients.

Theorem 4. *If Assumptions 1 and 2 are satisfied, then $\widehat{\beta}$ is a consistent estimator, that is:*

$$\widehat{\beta} \to^p \beta_0 \quad as \;\; T \to \infty$$

Theorem 4 shows that under standard technical assumptions, as long as the number of estimated VAR lags and the number of matched IRF coefficients do not increase too quickly,

[9] In particular, Lutkepohl (1988) and Lutkepohl and Poskitt (1991).

[10] This result is stated formally in the appendix as Lemma 6, since we use it directly in our proof of Theorem 1.

the common approach remains consistent. In the following section, we discuss the case when Assumption 2(iii) does not hold, and the number of matched IRFs is large relative to the number of estimated VAR lags, as is commonly the case in practice.

2.4 Bias Matters Asymptotically: An Example

The previous section shows that the common IRFME continues to be consistent when the data is from a VAR(∞) and the econometrician wishes to match IRFs at increasingly long horizons as the sample size grows. However, this result requires that the horizon of impulse responses matched, q, coincides with the order of the estimated VAR, r. This is usually not true in applications. In our primary empirical example, ACEL estimates a VAR with 4 lags and match impulse responses on horizons up to 20.

In this section, we show that in a realistic example where the number of matched IRFs grows faster than the degree of the estimated VAR, the common approach is inconsistent. Intuitively, this result stems from two causes. First, the mapping from VAR to IRF coefficients is nonlinear, and the degree of non-linearity increases with the horizon. This non-linearity leads to a bias in the (first-stage) estimation of IRFs, especially at longer horizons. Second, a standard "efficient" IRFME places more weight on those IRF coefficients with smaller estimated standard errors, which in practice are computed by the delta method under the assumption that first-stage IRF estimates are asymptotically normal. We show that standard errors computed in this way become very narrow on long horizons, and as a result, the IRFME is influenced disproportionately by the biased IRF coefficients.

2.4.1 Example

Assume that we have a sample from a univariate AR(1) process:

$$y_t = \beta_0 y_{t-1} + e_t, \quad e_t \sim N(0,1), \quad t = 1, ..., T.$$

We choose to work with a univariate process in this example for simplicity and leave a multivariate extension of our results for future work. Assume that one ignores the knowledge of the exact model and estimates an AR(r) process using OLS, namely

$$y_t = \sum_{j=1}^{r} A_{j,r} y_{t-j} + e_t. \tag{2.4.1}$$

The true coefficients in this case are $A_{1,r} = \beta$ and $A_{j,r} = 0, j \geq 2$. Although the econometrician includes more lags than are present in the true model, the estimated model of this example is correctly specified. This is in contrast to the sieve-VAR approach of the previous section where too few lags are included in the regression, and therefore, the model is misspecified.

Let $\left[\widehat{A}_{1,r}, \ldots, \widehat{A}_{r,r}\right]$ be the OLS coefficient estimates for the employed AR(r) model. Let $\widehat{\Theta}_{j,r}$ denote the estimated impulse response coefficient at horizon j obtained by inverting the estimated AR(r) process, as in the previous section. For a given value of β, the impulse response implied by the model is given by $\Theta_j(\beta) = \beta^j$.

As before, the matching estimator is computed as

$$\widehat{\beta} = argmin_{\beta} \sum_{j=1}^{q} \frac{(\widehat{\Theta}_{j,r} - \Theta_j(\beta))^2}{\omega_{j,r}}. \tag{2.4.2}$$

where the weights $\omega_{j,r}$ are equal to an estimate of the variance of $\widehat{\Theta}_{j,r}$. In keeping with common practice in empirical work, we assume that $\omega_{j,r}$ are computed by the delta method as

$$\omega_{j,r} = \sum_{k=1,m=1}^{r} \frac{\partial f_{j,r}(A_1, ..., A_r)}{\partial A_k} E(\hat{A}_{k,r} - A_k)(\hat{A}_{m,r} - A_m) \frac{\partial f_{j,r}(A_1, ..., A_r)}{\partial A_m}$$

where $f_{j,r}$ is the function that transforms a set of AR(r) coefficients into the impulse response at horizon j, in particular, $\widehat{\Theta}_{j,r} = f_{j,r}(\widehat{A}_{1,r}, ..., \widehat{A}_{r,r})$.

We make the following assumption about the rates of growth of r and q.

Assumption 3 Assume that $r_T < q_T < T$ all increase to infinity in such a way that

$$\frac{r_T^3}{T} \to 0, \qquad \frac{r_T}{\log(T)} \to \infty, \quad \frac{q_T^2}{T} \to \infty.$$

Most importantly, we assume that the horizon of matched IRFs increases faster than in Assumption 2(iii) of the previous section. We maintain the same upper bound on the rate of growth of r as before, but here we introduce a lower bound to ensure it does not grow too slowly.

The first order condition for the optimization problem (2.4.2) is

$$\sum_{j=1}^{q} \frac{(\widehat{\Theta}_{j,r} - \Theta_j(\hat{\beta}))}{\omega_{j,r}} \frac{\partial \Theta_j(\hat{\beta})}{\partial \beta} = 0. \tag{2.4.3}$$

The following theorem shows that in our setting when the horizon of matched impulse responses, q, increases too fast the first order condition at the true parameter value is not satisfied asymptotically, and thus the common approach is inconsistent.

Theorem 5. *If assumption 3 is satisfied in the described estimation setting, then there exists* $\varepsilon > 0$ *such that*

$$\lim_{T\to\infty} P\left\{\left|\sum_{j=1}^{q_T} \frac{(\widehat{\Theta}_{j,r_T} - \Theta_j(\beta_0))}{\omega_{j,r_T}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta}\right| > \varepsilon\right\} = 1. \tag{2.4.4}$$

Hence, under the realistic condition that $r_T \to \infty$, the consistency of the common approach depends crucially on the rate of growth of q_T.

The proof of Theorem 5 focuses mainly on the comparison between the bias of the first-stage IRF estimates, and the weights used in the objective function. Two potential sources of first-stage bias are present. First, IRF coefficients are nonlinear functions of AR coefficients. Second, estimates of AR coefficients are biased because the residuals do not satisfy strict exogeneity. In practical settings, a third source of bias arises from the misspecification of the estimated model. That is, there is a bias due to the fact that the econometrician estimates a regression with only a finite number of lags, but the true model has an infinite number of lags. This third type of bias may be very important, especially when the process has any roots that are close to unity. For simplicity, the estimated model in our example is correctly specified, but one can reasonably expect that estimating a misspecified model would not weaken the conditions required to obtain consistency of the common approach. In proof of Theorem 5, we find that the first type of bias is the most relevant for our current example, especially for IRF coefficients at further horizons.

As mentioned previously, the objective function weights are equal to standard estimates of the variance of the first-stage IRF coefficient estimators. The particular variance estimator chosen is the one typically used in practice and is based on the delta method. Under standard asymptotics where r and q are fixed, this choice of weights would lead to an efficient estimator $\widehat{\beta}$. However, when q is allowed to grow quickly enough, weights based on standard asymptotics increase rapidly with the horizon, as the given IRF coefficient variance estimates decline with the horizon. As a result, large weight is placed on the relatively more biased long-horizon IRF coefficient estimates, and the objective function is not minimized at β_0.

Given that the consistency of the common approach breaks down when q is too large, we turn now to a comparison with the alternative bias corrected approach. In the following sections, we present simulation evidence that the bias corrected approach improves the properties of IRFMEs when q is large.

2.4.2 Simulation Results

In order to support the findings of the prior section and demonstrate that the bias corrected approach is an improvement, we ran simulations of the described example. Specifically, we ran 500 simulations with $T = 200$, $M = 200$, $\beta_0 = 0.95$, and $r = 5$. The weighting matrix we use is numerically identical for both estimators, and it is computed according to the formula of the prior section.

Table 1: AR(1) Simulation Results

Horizon (q):	5	10	15	20
Bias				
Common	-0.0100	-0.0108	-0.0115	-0.0118
Bias Corrected	-0.0009	-0.0017	-0.0028	-0.0038
Mean Squared Error				
Common	0.00144	0.00091	0.00083	0.00083
Bias Corrected	0.00127	0.00078	0.00071	0.00071
Variance				
Common	0.00135	0.00080	0.00070	0.00069
Bias Corrected	0.00127	0.00078	0.00070	0.00070

Notes: $N = 200$, $M = 200$, $\beta_0 = 0.95$, $r = 5$.

Table 1 reports the average bias, mean squared error, and variance for each estimator across the simulated samples. As expected, the bias corrected estimator acheives a lower average bias. At the same time, the variances of the two estimators are very similar, and as a result, the average MSE is lower for the bias corrected estimator. These results confirm that the common approach is biased in this setting, and it is outperformed by the bias corrected approach.

2.5 Example: Altig et al (2011)

In order to further compare the bias corrected approach with the common approach in practical settings, we use the model of ACEL as an example. As described in the introduction, ACEL is a relevant example for us to consider because it uses the common IRFME with a q that is large relative to T and r. Specifically, the benchmark estimation of ACEL is done with $q = 20$, $r = 4$, and $T = 170$ where they match the responses of 10 macroeconomic variables to the 3 shocks from their model. First, we discuss the key features of the model. Second, we show the results of simulations comparing the performance of the two versions of the IRF matching estimator. To conclude the section, we discuss how the estimation results change when applying the bias correction method.

2.5.1 The Model

The primary goal of ACEL is to reconcile the conflict between micro- and macro-based evidence on the frequency at which firms adjust their prices. On one hand, recent micro-level data indicate that firms change prices roughly every 1.5 quarters, as argued by Bils and Klenow (2004), Golosov and Lucas (2007), and Klenow and Kryvtsov (2008). In contrast, aggregate inflation is persistent. As a result, previous calibrated macro models that match the inflation persistence require firms that update prices no more frequently than every 5 quarters.[11] ACEL attempts to explain these facts by introducing firm-specific capital into the homogeneous capital model of Christiano et al (2005).

Here, we quickly summarize the setup of the model to make interpretation more clear for the results that follow. The model contains three aggregate shocks that drive its dynamics. These are a monetary policy growth rate shock, ε_M, a neutral technology growth rate shock, ε_z, and a capital embodied technology growth rate shock, ε_Υ. The three shocks are assumed to be uncorrelated over time and uncorrelated with each other. In the homogeneous capital version of the model, there are households, a government, a representative, perfectly competitive goods-producing firm, and monopolistic intermediate goods firms.

[11]See, for example, Smets and Wouters (2003), Rabanal and Rubio-Ramirez (2005), and Eichenbaum and Fisher (2007).

Final goods production is given by

$$Y_t = \left[\int_0^1 y_t(i)^{\frac{1}{\lambda_f}} \, di \right]^{\lambda_f}$$

The price of the final good is denoted by $P_{t.}$.

There are a continuum of intermediate goods used to produce the final good, and intermediate goods firms face a fixed cost in production. Production of intermediate goods firms is given by

$$y_t(i) = \begin{cases} K_t(i)^{\alpha} (z_t h_t(i))^{1-\alpha} - \phi z_t^* & \text{if } K_t(i)^{\alpha} (z_t h_t(i))^{1-\alpha} \geq \phi z_t^* \\ 0 & \text{otherwise} \end{cases}$$

where $i \in (0,1)$ indexes the intermediate good, K_t is capital, and h_t is labor. The shock ε_z is a growth rate shock to z_t, and ε_Υ is a growth rate shock to z_t^*. Intermediate goods firms are risk neutral, and they rent capital and pay wages. The markets for labor and capital are perfectly competitive. Intermediate goods prices are denoted by $P_t(i)$. Each period, intermediate goods firms are allowed to alter $P_t(i)$ with probability $1 - \xi_p$, and with probability ξ_p their price is indexed to inflation as $P_t(i) = \pi_{t-1} P_{t-1}(i)$. In each period, intermediate goods firms first observe $\varepsilon_{z,t}$ and $\varepsilon_{\Upsilon,t}$, then set prices $P_t(i)$, then observe a monetary policy shock and demand for goods, and finally choose labor and capital for production. In the firm-specific capital version of the model, intermediate goods firms own capital and make independent investment decisions. The sequence of events is the same except that after setting prices and before observing the monetary policy shock, the intermediate goods firms choose how much to invest and current period capital utilization.

Households supply labor and capital to firms, invest in the creation of future capital, and consume the final good. The government sets monetary policy in response to the two technology shocks. Further details about the behavior of households and the government is not useful to understand what follows, so we do not describe them further here.

The equilibrium conditions of the two models are the same except for one that governs inflation dynamics. In both models, inflation dynamics are given by

$$\Delta\hat{\pi}_t = E\left[\beta\Delta\hat{\pi}_{t+1} + \gamma\hat{s}_t|\Omega_t\right] \tag{1}$$

where $\hat{\pi}$ is the growth rate of inflation, β is the household's discount factor, $\hat{s}_t$ is the average marginal cost of production in terms of the final good, Ω_t is the information available to firms when they set their prices, and γ is a reduced form parameter. The mapping from the structural parameters to γ is where the two models differ. In the equilibrium of both models, we can write

$$\gamma = \frac{(1-\xi_p)(1-\beta\xi_p)}{\xi_p}\chi. \tag{2}$$

In the homogenous capital model $\chi = 1$, while in the firm specific capital model χ is defined implicitly in terms of other structural parameters. The average time it takes for firms to reoptimize is given by $(1-\xi_p)^{-1}$.

The inability of the homogeneous capital model to simultaneously match micro and aggregate price data can be seen from these two equations. Take $\beta = 1.03^{-.25}$ to be consistent with a 3% annual real interest rate in equilibrium. When ACEL estimate (1) on aggregate data, they obtain a low value of $\gamma \approx 0.04$,[12] which is consistent with previous estimates, such as in Eichenbaum and Fisher (2007). Plugging this value into the homogeneous capital model and solving (2), gives $\xi_p = 0.82$, which implies that firms reoptimize prices on average every 5.6 quarters. Conversely, if we assume that firms reoptimize prices every 1.5 quarters to be consistent with the micro-level evidence, solving (1)-(2) gives $\xi_p = 0.33$ and $\gamma = 1.34$. Hence, it is infeasible for the homogeneous capital model to capture both micro and macro price data. In the firm specific capital model, however, χ varies with other model parameters, so

[12] Altig et al (2011) estimate $\gamma = 0.014$ for post-1982 data, which only worsens the fit of the homogeneous capital model.

in principle, it could match the price data depending on the full set of parameter estimates.

2.5.2 Estimation Methodology and Data

As previously discussed, ACEL estimates their model by matching VAR estimates of the IRFs of 11 macroeconomic indicators to the three model shocks with estimates of the corresponding true IRFs as implied by the model for given parameters. In order to compute model dynamics, they solve for the state-space representation and establish a mapping between model variables and those observed in the data. This allows them to directly compute the true response of observables to the three shocks, assuming the model is accurate. These model IRFs correspond to $\theta(\Gamma)$ in the notation above. To compute $\hat{\theta}(Y_T)$, they estimate a VAR(r) on the data and impose assumptions that are consistent with the model to identify the model shocks. Specifically, the identifying restrictions they use are that only technology shocks affect long run labor productivity and only embodied technology shocks affect the long run relative price of investment with respect to consumption. The structural IRF estimates are computed using the instrumental variables approach of Watson and Shapiro (1988).

When computing the IRFME, ACEL treats γ as a structural parameter. Since the two models only differ in their mapping from γ into the structural parameters, this allows both models to be estimated simultaneously, thereby reducing computational complexity. The benchmark estimation of ACEL uses $r = 4$, $q = 20$. They also hold fixed 19 of the model parameters at values that are consistent with the literature, and estimate the remaining 18 parameters. Finally, the weighting matrix W is diagonal, and its elements are estimates of the inverse of the variance of the corresponding first-stage IRF coefficients.

The data is taken from the DRI Basic Economics Database, except the price of investment is taken from Fisher (2002) and monetary transactions balances is taken from the online database of the Federal Reserve Bank of St. Louis. The variables are quarterly measures of output, monetary transactions balances growth, inflation, Federal Funds rate, capacity utilization, average hours, real wage, consumption, investment, velocity, price of investment, and total money growth.[13] The data are reported at a quarterly frequency for the period 1959Q2 - 2001Q4.

2.5.3 Simulations

As a means to compare the properties of the common IRFME with the bias corrected version described above, we perform simulations on the ACEL model. However, this is complicated by the fact that there are only 3 shocks in the model, and the goal is to match the impulse responses of 11 variables. Since the model is presumed to explain only a part of the actual DGP, simulating a sample from it involves the computation of two components. In particular, the vector of variables whose responses we wish to match can be written as $Y_t = Y_t^{Other} + Y_t^{Identified}$ where Y_t is the observed data, $Y_t^{Identified}$ is the portion of the data that is generated from the ACEL model, and Y_t^{Other} is the remainder. This follows from the Wold representation of the data,

$$Y_t = (I - B(L))^{-1} C\varepsilon_t$$

where ε_t are the structural shocks. Let ε_{1t} consist of the three shocks in the model and C_1 consist of the three columns of the matrix C that correspond to these shocks. Let ε_{2t} be the remainder of the shocks in ε_t and C_2 be the remainder of the columns of C. Using this

[13] See ACEL for details on how these series were computed.

notation, we can rewrite the previous expression as

$$Y_t = Y_t^{Other} + Y_t^{Identified}$$
$$Y_t^{Identified} = (I - B(L))^{-1} C_1 \varepsilon_{1t}$$
$$Y_t^{Other} = (I - B(L))^{-1} C_2 \varepsilon_{2t}$$

Simulating $Y_t^{Identified}$ for given model parameters is done by simulating a sequence of shocks and plugging these into the state-space representation. Following the authors' proposed procedure, we simulate Y_t^{Other} parametrically from the VAR coefficients that were estimated on the actual data.[14] More precisely we generate the "other" component as

$$Y_t^{Other} = \hat{B}(L) Y_t^{Other} + \hat{C}_2 \hat{\varepsilon}_{2t}$$

where $\hat{C}$ and $\hat{B}(L)$ are taken from the estimates on the data and $\hat{\varepsilon}_{2t}$ are drawn from $N\left(0, \hat{\Sigma}_{22}\right)$ where $\hat{\Sigma}_{22}$ is a consistent estimator of the variance covariance matrix for ε_{2t} and was estimated from the data. This procedure requires initial values for Y_t^{Other} which we set equal to the initial observations in the data. This choice for Y_t^{Other} is designed to make the simulated samples as similar to the actual data as possible.[15]

2.5.4 First-Stage Bias vs. Identification

Having described the model and how to simulate data from it, we now turn to a comparison of the two IRFMEs via simulations. The results of the prior section show that first-stage bias can lead to a breakdown in the common approach when q is large relative to r and T. As a preliminary illustration of how important this problem is for the ACEL model, we com-

[14]See Altig et al (2004) for further details.

[15]The bias corrected estimator requires the computation of $X_T(\Gamma) = X_T^{Other} + X_T^{Identified}(\Gamma)$ where X_T^{Other} is computed as described for Y_T^{Other}. For estimation on simulated samples, we reestimate $\hat{C}, \hat{B}(L)$ and $\hat{\Sigma}$ from the simulated sample to generate X_T^{Other}.

pare the size of the first-stage bias to the degree of identification of γ. Bias is illustrated by comparing the model implied true IRFs with the average of those estimated from simulated samples for the same parameter values. The degree of identification of γ is illustrated by comparing the model implied IRFs across different values of γ.

Figure 1 illustrates the comparison of first-stage bias and the degree of identification of γ for the response of output to the three model shocks. In particular, it plots the model implied IRFs against the average of those estimated from simulated samples for $\gamma = 0.04$, the ACEL benchmark estimate of that parameter and a value consistent with the aggregate data. These two series are labeled "Model ($\gamma = 0.04$)" and "Bias Corrected" respectively. The plot also shows the model implied IRFs for $\gamma = 1$, a value that is more in line with the micro data for the homogeneous capital model.[16] This series is labeled "Model ($\gamma = 1$)." Finally, for comparison, we include the impulse responses estimated via VAR on the actual data, labeled "Data," along with their confidence intervals. We chose output as the response variable for this figure simply because it is the most important and commonly used measure of aggregate activity.

The first noteworthy feature of Figure 1 is that the degree of identification of γ is small. This can be seen from the difference between the model's impulse responses for the two values of γ, as compared with the difference between the responses estimated from the data and the model's at the ACEL estimate of $\gamma = 0.04$. Strikingly, the former difference is consistently smaller across shocks and horizons. In other words, the fit of the model for the responses of output is little changed when making large changes to γ. Furthermore, this contrasts with ACEL's reported standard errors of 0.02 for γ. The model's IRFs for both of these parameter values are inside the reported confidence bands for the IRFs estimated from data, despite the fact that the latter is way outside of the reported confidence interval for $\hat{\gamma}$. Consistent with our results of Section 4, this suggests that the distribution of $\hat{\gamma}$ is not be

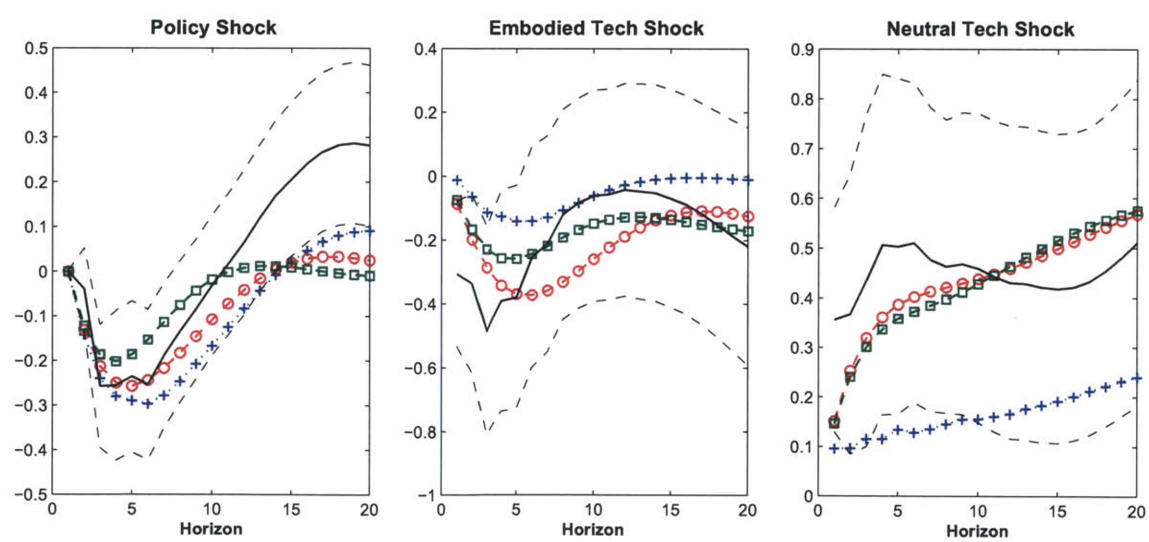
[16] All parameters other than γ are set to the ACEL benchmark estimates for all series in Figure 4.

well approximated by a normal distribution.

Second, the bias in the estimated IRFs is large relative to the degree of identification of γ. The size of the bias can be seen as the difference between the series "Model ($\gamma = 0.04$)" and "Bias Corrected." Across horizons and shocks, the bias is at least as large as the degree of identification of γ, and for the two technology shocks, it is strictly larger. The discrepancy is especially noticeable for the neutral technology shock where the size of the bias is roughly equal to the radius of the confidence interval for the estimated responses and the model's responses are nearly identical for the two values of γ. Taken together, this evidence suggests that correcting for first-stage bias is very important for evaluating the performance of the model. Figure A1 in the appendix shows a similar plot for the responses of all 10 variables, and confirms that the size of the bias is large in many cases, relative to the fit of the model.

Figure 1: Response of Output to Model Shocks

(Data - Solid, Model ($\gamma = 0.04$) - O, Model ($\gamma = 1$) - □, Bias Corrected - +)



Notes: Parameter values other than γ are from ACEL benchmark estimates. Dashed lines show the 95% confidence intervals for the IRFs estimated from the data.

2.5.5 Simulation Results

In this section, we use simulations to compare the properties of the common and bias corrected approaches for the ACEL model. That is, we simulate many samples from the ACEL model. For each sample, we estimate the key parameter, γ_0, by both the common approach and the bias corrected approach, while treating all other parameters as known. Then, we compute the mean bias, mean squared error, and variance for each estimator across the simulated samples. Finally, we compare the performance of the two approaches to see first whether the common approach shows significant bias, as would be consistent with our prior theoretical results, and second, whether the bias corrected approach provides an improvement.

For our baseline simulations, we set $\gamma_0 = 1$, a value that is well outside of ACEL's estimated confidence interval and one which makes the homogenous capital model roughly consistent with micro-level evidence.[17] We chose this value of γ_0 first because it is away from the zero lower-bound on the parameter space of γ. It is also motivated by the findings in our previous section, namely that the degree of identification between $\gamma = 1$ and γ near zero is small, relative to the magnitude of the first-stage bias. In other words, we want to see whether first-stage bias leads the common approach to produce $\hat{\gamma}$ near zero even when γ_0 is not and whether the bias corrected approach alleviates this problem.

For other settings in the baseline case, we use the same approach as in ACEL. That is, we set $r = 4$, $q = 20$, $T = 170$, and we match the IRFs of all 10 variables to all three model shocks. For the bias corrected estimator, we set $M = 100$. The weighting matrix we use is the same as in ACEL and identical for both estimators. Also for both estimators, numerical optimization is done by grid search with boundaries on the parameter space set at 0 and 2 with an interval length of 0.05. We run 100 simulations to compare the two estimators.

The simulation results are shown in Table 2. Strikingly, even with $\gamma_0 = 1$, the average

[17] All other parameters are set at the ACEL benchmark estimates.

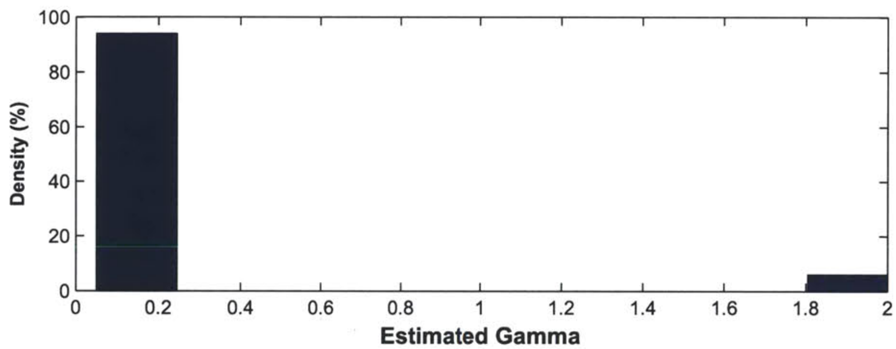
common approach estimate was only 0.17, closer to the benchmark estimate reported in ACEL. In contrast, the average bias corrected estimate was 0.85, closer to the true value. Thus, the common approach had a bias that was almost 6 times larger than that of the bias corrected approach. As a result, even though the variance of the bias corrected approach was larger, its mean squared error was more than 25% lower.

Figure 2 plots the histogram for each estimator. It shows several interesting features. Starting with Panel A, almost every common approach estimate was under 0.20. Hence, even if the true value of γ is far from 0, it is not surprising that ACEL estimated it at $\hat{\gamma} = 0.04$ when using the common approach. The handful of estimates that were not close to 0 were close to 2, the other boundary, and in particular, no estimates were in a reasonable neighborhood of the true value. This sheds doubt on the efficacy of the common approach in this setting. It is severely biased towards 0, and its distribution is far from normal, causing the usual standard errors to be unreliable. In Panel B, we see that the bias corrected estimator does better, with at least some estimates near the true parameter value, though its distribution is still not close to normal.

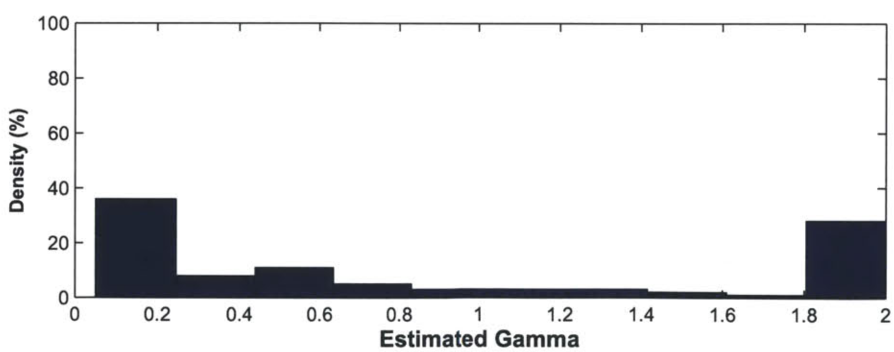
The remainder of Table 2 reports robustness checks for our baseline simulations. Panel B reports simulations where the IRFMEs match only the IRFs to the policy shock. Both estimators did worse in bias and MSE than the baseline case, but the bias corrected version continued to outperform the common approach. In fact, the common estimator produced $\hat{\gamma} = 0.05$ in every single sample for this case, which corresponds to the largest possible bias, since $\gamma_0 = 1$. Panel C shows the results when matching only the IRFs to the embodied technology shock. This is the only case where the common estimator had lower bias, however, it still underperformed in terms of MSE. In Panel D, we show the results when only matching the IRFs to the neutral technology shock. Again the bias corrected approach outperformed in terms of bias, though the two were comparable in terms of MSE. According to simulations reported in ACEL, the common approach performed better when the first-stage estimates are based on a VAR(6) (i.e., $r = 6$). We include this case in Panel E and find that the results are much the same as in the baseline. The last scenario we consider is only matching 15 steps of the IRFs, rather than 20. The outcome is reported in Panel F. Even though the number of first stage parameters are reduced, the results are close to the baseline. Overall, for the ACEL setting, we find strong evidence that the common IRF estimator has large bias, and it is outperformed by the bias corrected approach in terms of both MSE and bias.

Figure 2: ACEL Simulation Results

Panel A: Common Approach



Panel B: Bias Corrected Approach



Notes: Parameter values are from ACEL benchmark estimates, except we set $\gamma = 1$. Results are based on 100 simulations. Optimization is done by grid search.

Table 2: Simulation Summary Statistics			
Approach	**Mean Bias**	**Mean Squared Error**	**Variance**
Panel A: Baseline			
Common	-0.83	0.91	0.22
Bias Corrected	-0.15	0.67	0.65
Panel B: Only Policy Shock			
Common	-0.95	0.90	0.00
Bias Corrected	-0.42	0.76	0.59
Panel C: Only Embodied Tech. Shock			
Common	-0.11	0.94	0.94
Bias Corrected	-0.27	0.79	0.73
Panel D: Only Neutral Tech. Shock			
Common	-0.65	0.80	0.37
Bias Corrected	-0.02	0.80	0.81
Panel E: $r = 6$			
Common	-0.89	0.90	0.11
Bias Corrected	-0.62	0.78	0.40
Panel F: $q = 15$			
Common	-0.90	0.90	0.09
Bias Corrected	-0.12	0.79	0.78

Notes: Only γ is estimated, and $\gamma_0 = 1$. All other parameters are treated as known and set at the benchmark estimates from ACEL.

2.5.6 Estimation Results

We have shown that the common approach produces biased and non-normal estimates in the practical setting of ACEL. Furthermore, the bias corrected approach provides an improvement in our simulations. In this section, we discuss how the parameter estimates for the model change when using the bias corrected approach. We set $M = 200$, and estimate the model parameters using the same settings as in the benchmark case of ACEL.

The key parameters estimated by the bias correction approach are reported in Table 3 alongside the original estimates from ACEL. For the baseline case, the bias corrected estimates are substantially different from those obtained using the common approach. Most importantly, we obtain $\hat{\gamma} = 0.83$, which puts the model at odds with the evidence on aggregate inflation that γ is close to zero. The bias corrected estimates imply that in the homogeneous capital model, firms reoptimize prices every 1.71 quarters, while in the firm-specific capital model, they do so every 1.1 quarters. According to our estimates, it is not clear that the firm-specific capital model does any better than the homogeneous capital model at explaining the micro evidence on price behavior, and neither is it able to match the macro evidence. These findings contradict the primary conclusions of ACEL and show that accounting for first-stage bias can have a meaningful impact on estimation results in such settings. Table A1 in the appendix shows the full set of parameter estimates, and confirms that many parameter estimates change when using the bias corrected approach. For example, the inverse of the elasticity of investment with respect to a temporary change in the price of capital, S'', drops from 3.28 under the common approach to 1.00.

The third and fourth columns of Table 3 report robustness checks for the bias corrected estimates. In the third column, we set the weighting matrix, W, equal to the identity. If the first-stage IRF coefficient variance estimates are badly biased, then we would expect this change to significantly affect our estimates. Our theoretical results also indicate that weighting may play an important role in the properties of the estimators. However, we find the resulting estimate of γ and implied frequency of price adjustment are similar to the baseline case. In the fourth column, we consider the bias corrected approach where IRFs are matched only up to a horizon of 15 quarters. According to our theoretical results, reducing the horizon of IRFs being matched could also help reduce the bias. However, doing this increases the estimate of γ and makes the homogeneous and firm-specific capital models even more similar in terms of their predictions about the frequency that firms adjust prices. Thus, we conclude the correcting for first-stage bias is important in the setting considered by ACEL and doing so actually reverses some of that paper's main conclusions.

Table 3: ACEL Key Parameter Estimates

Approach:	**Common**	**Bias Corrected**			
	Baseline	**Baseline**	$W = I$	$q = 15$	**Alternate**
γ	0.04	0.83	0.89	1.20	0.01
Homogeneous Capital:					
ξ_p	0.82	0.41	0.42	0.35	0.92
Price Adj. Freq. (quarters)	5.60	1.71	1.71	1.53	13.09
Firm-Specific Capital:					
ξ_p	0.34	0.09	0.28	0.35	0.55
Price Adj. Freq. (quarters)	1.51	1.10	1.38	1.54	2.22

2.5.7 Evidence on Precision of Original Estimates

Finally, we consider confidence intervals for γ. In our simulations, we showed that both the common and bias corrected IRFMEs produce estimators with non-normal distributions. Therefore, inference based on typical standard errors is likely to be misleading. Instead, we use simulations to construct our own confidence intervals for γ that do not rely on the approximation that $\hat{\gamma}$ is normally distributed. We compute critical values via a parametric bootstrap as follows:

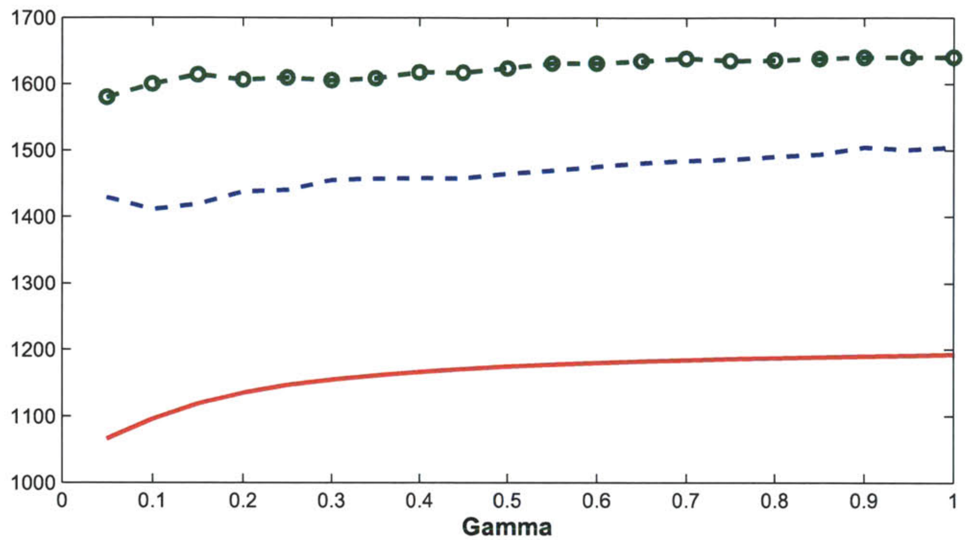
- Form a grid for γ. We use a grid with boundaries at 0 and 1 and an interval of 0.05;
- For each value of γ in the grid, simulate B samples from the model with sample size equal to that of the actual data (i.e., $T = 170$);
- For each simulated sample, compute the VAR-based estimates of the IRFs, $\hat{\theta}(Y_T)$, and weighting matrix, W, as in ACEL;
- For each simulated sample, compute the value of the objective function using the quantities from the previous step;
- For each value of γ, record the 95% and 90% critical values of the objective function;
- Plot these critical values against the objective function computed on the actual data as a function of γ.

First, we compute a bootstrap confidence interval for the common approach. The plot described in the final item for the common approach is displayed in Panel A of Figure 3. The confidence interval contains all values of γ where the objective function is below the critical value for the chosen significance level. The figure shows that the 90% confidence interval for γ under the common approach includes the entire interval between 0 and 1. This starkly contrasts ACEL's reported 95% confidence interval of [0, 0.08]. Panel B shows the same plot for the bias corrected estimator, holding the parameters at the common approach estimates. In this case, the 95% confidence interval for γ would only include the range [0, 0.10]. It is important to note that this is not a valid confidence interval because the parameters are held at the estimates from the common approach, not the bias corrected estimates. Still, Panel B illustrates that inference can be quite different for the two estimators in realistic settings.

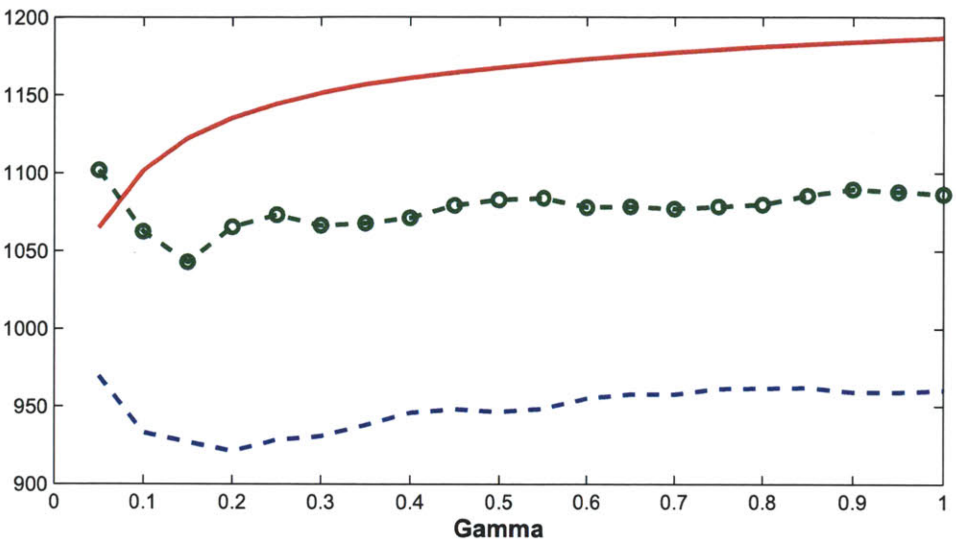
Figure 3: Confidence Intervals

(Objective - Solid, 90% CV - Dash, 95% CV - Circle)

Panel A: Common Approach



Panel B: Bias Corrected Approach

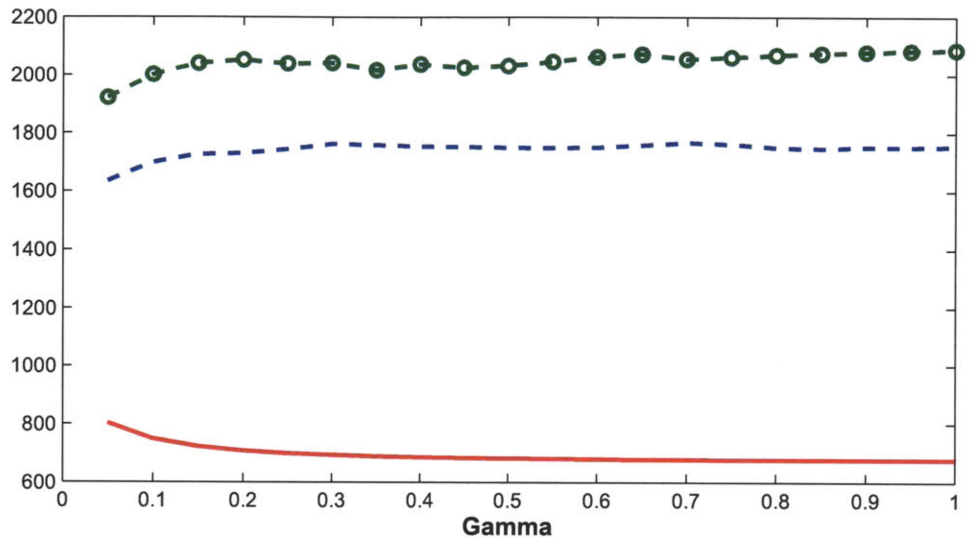


Notes: Parameter values are from ACEL benchmark estimates and $B = 1,000$.

Next, we turn to constructing confidence intervals for γ based on the bias corrected approach. The analog of Panel A of Figure 3 for the bias corrected approach is plotted in Figure 4. As with the common approach, we find that the confidence interval for γ includes all values between 0 and 1. Thus, standard confidence intervals based on normality of the parameter estimates are not accurate in a realistic setting where q is large relative to r and T.

Figure 4: Confidence Intervals

(Objective - Solid, 90% CV - Dash, 95% CV - Circle)



Notes: Parameter values are from the Bias Corrected baseline estimates in Table 3, and $B = 1,000$.

In order to emphasize the implications of our large confidence intervals, we compare our estimates with another set of parameters that produce a similar value of the objective function. These parameter values are listed in the last column of Table A1. The alternate parameters generate an objective value of 737, which is not significantly more than the baseline of 694. However, the alternate parameters have entirely different economic implications from the baseline. First, $\hat{\gamma} = 0.01$ which is small enough to be consistent with micro evidence. Moreover the set of parameters implies that under the homogeneous capital model firms would take 13.1 quarters to reoptimize, while in the firm specific capital model firms would update prices every 2.2 quarters. These numbers would make the latter roughly consistent with actual firm behavior while the former would do poorly in that respect. The estimates of S'' and σ_a also change leading to much different predictions on those dimensions. Therefore, we conclude that the model parameters are not precisely estimated in terms of either parameter values or economic implications.

2.6 Conclusion

In this paper, we have considered the properties of the common impulse response function matching estimator ("IRFME") in settings with many parameters. Specifically, we extended existing theoretical results to the case where the true model is an infinite order VAR, and the number of impulse responses being matched is non-negligible relative to the sample size. First, we proved that the common approach remains consistent and asymptotically normal as long as the horizon of IRFs being matched grows slowly enough. However, we also provided a simple example where the number of matched IRFs grows too quickly, and the common approach is not consistent. Finally, we used simulations to evaluate the performance of the common IRFME in the setting considered recently by ACEL, and we compared it with a bias corrected approach. In this realistic example, we found that the common estimator has a large bias, and a distribution that is far from normal. Meanwhile, the bias corrected estimator performs better, and using it reverses ACEL's main conclusions. Our findings suggest that the common IRFME performs poorly in situations where the sample size is not much larger than the horizon of IRFs being matched, and in those situations, the bias corrected approach with bootstrapped standard errors performs better.

2.7 Appendix

2.7.1 Proofs

In the following proofs, we consider two norms: $\|B\|^2 = trace(B'B)$ and $\|B\|_1^2 = max_x \frac{x'B'Bx}{x'x}$. We use the following known inequalities: $\|AB\|^2 \leq \|A\|^2\|B\|_1^2$, $\|AB\|^2 \leq \|A\|_1^2\|B\|^2$, $\|AB\|_1^2 \leq \|A\|_1^2\|B\|_1^2$. In what follows, C will denote a generic constant.

2.7.1.1 Proof of Theorem 1 (Positive Result)

We will use the following result on the asymptotic normality of infinitely many impulse responses in sieve-VAR estimation:

Lemma 6. [18] *Let Assumption 1 be satisfied, and let λ_r denote a sequence of $K^2 r \times 1$ vectors (here $r = r_T$) such that $0 < c_1 \leq \|\lambda_r\|^2 < c_2 < \infty$ then*

$$\sqrt{T}\lambda_r'(\widehat{\theta}_{r,r} - \theta_r)/s_r \Rightarrow N(0,1),$$

where $s_r = \lambda_r'\Omega_r\Lambda_r$, and $\Omega_r = [\Sigma^{-1} \otimes \sum_{j=0}^{n-1} \Theta_j \Sigma \Theta_{m-n+j}']_{n,m=1,..,r}$. In particular, the asymptotic variance of $\widehat{\theta}_{r,q}$ at a fixed horizon q is $\Sigma^{-1} \otimes \sum_{j=0}^{q-1} \Theta_j \Sigma \Theta_j'$ (and is increasing in h).

For the proof of Theorem 4, we need the following Lemma about the norm of infinitely many impulse responses.

Lemma 7. *Under Assumption 1, we have*

$$\|\widehat{\theta}_{r,r} - \theta_r(\beta_0)\|^2 = O_p(r/T)$$

[18] Lemma 6 is taken from Lutkepohl and Poskitt (1991).

Proof of Lemma 7 Here, we follow quite closely the proof of Theorem 2 in Lutkepohl and Poskitt (1996). In particular, we can write the relation between estimated impulse responses and the OLS VAR coefficients in the following way:

$$\hat{\theta}_{r,r} - \theta_r(\beta_0) = \hat{R}(r)\left[\hat{a}(r) - a(r)\right]$$

where $a(h) = vec\left[A_1(h), \ldots, A_h(h)\right]$,

$$\hat{R}(r) = \begin{bmatrix} E_r \otimes I_K \\ \cdots \\ \sum_{j=0}^{r-1} E_r \widehat{\mathbf{A}}_r^{\prime r-j-1} \otimes \Theta_j \end{bmatrix}, \qquad E_r = (I_K, \mathbf{0}, \ldots, \mathbf{0}).$$

The coefficients are estimated by OLS so that

$$\hat{a}(r) = vec\left[\left\{\frac{1}{T}\sum_{t=1}^{T} y_t Y_{t-1}(r)'\right\}\hat{\Gamma}_T(r)^{-1}\right] = \left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right) vec\left[\frac{1}{T}\sum_{t=1}^{T} y_t Y_{t-1}(r)'\right],$$

where $Y_t(r) = \left[y_t', \ldots, y_{t-r+1}'\right]'$, $X(r) = \left[Y_0(r), \ldots, Y_{T-1}(r)\right]$, $\hat{\Gamma}_T(r) = \frac{1}{T}X(r)X(r)'$.

Next, define

$$V_{1T} = vec\left[\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t Y_{t-1}(r)'\right]$$

and

$$V_{2T} = vec\left[\frac{1}{T}\sum_{t=1}^{T} \left[\varepsilon_t(r) - \varepsilon_t\right] Y_{t-1}(r)'\right],$$

where $\varepsilon_t(r) = y_t - \sum_{j=1}^{r} A_j y_{t-j}$ is the residual in the theoretical VAR cut after r lags, while$\varepsilon_t = y_t - \sum_{j=1}^{\infty} A_j y_{t-j}$ is the true error. We have

$$\begin{aligned} V_{1T} + V_{2T} &= vec\left[\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(r) Y_{t-1}(r)'\right] = vec\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{r} A_j y_{t-j}\right) Y_{t-1}(r)'\right] \\ &= vec\left[\frac{1}{T}\sum_{t=1}^{T} y_t Y_{t-1}(r)'\right] - vec\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{r} A_j y_{t-j}\right) Y_{t-1}(r)'\right]. \end{aligned}$$

This implies

$$\hat{a}(r) = \left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right)\left(V_{1T} + V_{2T} + vec\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{r} A_j y_{t-j} Y_{t-1}(r)'\right]\right)$$
$$= \left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right)(V_{1T} + V_{2T}) + vec\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{r} A_j y_{t-j} Y_{t-1}(r)' \hat{\Gamma}_T(r)^{-1}\right].$$

Looking at the rightmost piece of the last expression:

$$vec\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{r} A_j y_{t-j} Y_{t-1}(r)' \hat{\Gamma}_T(r)^{-1}\right]$$
$$= vec\left[a(r)' \sum_{t=1}^{T} Y_{t-1}(r) Y_{t-1}(r)' \left[X(r) X(r)'\right]^{-1}\right] = a(r).$$

Thus,

$$\hat{a}(r) - a(r) = \left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right)(V_{1T} + V_{2T}).$$

Then, we can rewrite the object of primary interest as

$$\theta_{r,r} - \theta_r(\beta_0) = \hat{R}(r)\left[\hat{a}(r) - a(r)\right]$$
$$= \hat{R}(r)\left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right)(V_{1T} + V_{2T})$$
$$= q_1(r) + q_2(r),$$

where

$$q_1(r) = R(r)\left(\Gamma(r)^{-1} \otimes I_K\right) V_{1T},$$
$$q_2(r) = R(r)\left(\Gamma(r)^{-1} \otimes I_K\right) V_{2T} + R(r)\left(\left[\hat{\Gamma}_T(r)^{-1} - \Gamma(r)^{-1}\right] \otimes I_K\right)(V_{1T} + V_{2T})$$
$$+ \left(\hat{R}(r) - R(r)\right)\left(\hat{\Gamma}_T(r)^{-1} \otimes I_K\right)(V_{1T} + V_{2T}).$$

Standard inequalities imply

$$
\begin{aligned}
\|q_1(r)\| &\leq \|R(r)\| \left\|\Gamma(r)^{-1}\right\|_1 \|V_{1T}\|, \\
\|q_2(r)\| &\leq \|R(r)\| \left\|\Gamma(r)^{-1}\right\|_1 \|V_{2T}\| + \|R(r)\| \left\|\hat{\Gamma}_T(r)^{-1} - \Gamma(r)^{-1}\right\|_1 (\|V_{1T}\| + \|V_{2T}\|) \\
&\quad + \left\|\hat{R}(r) - R(r)\right\| \left\|\hat{\Gamma}_T(r)^{-1}\right\|_1 (\|V_{1T}\| + \|V_{2T}\|).
\end{aligned}
$$

Lutkepohl and Poskitt (1996) show that the following statements are implications of Assumption 1:

$$
\begin{aligned}
\|R(r)\| \leq \sum_{i=0}^{r} \sum_{j=1}^{i-1} \|A_i\|^{r-j+1} \|\Theta_j\| < \infty, \\
\sqrt{r} \left\|\hat{R}(r) - R(r)\right\| = o_p(1), \\
\|V_{1T}\| = O_p\left(\sqrt{\frac{r}{T}}\right), \|V_{2T}\| = o_p\left(\sqrt{\frac{1}{T}}\right), \\
\sqrt{r} \left\|\hat{\Gamma}_T(r)^{-1} - \Gamma(r)^{-1}\right\| = o_p(1), \\
\left\|\Gamma(r)^{-1}\right\| < \infty, \left\|\hat{\Gamma}_T(r)^{-1}\right\| < \infty.
\end{aligned}
$$

Using all of the inequalities above, we are led to the conclusion

$$
\|q_1(r)\|^2 = O_p\left(\frac{r}{T}\right), \qquad \|q_2(r)\|^2 = o_p\left(\frac{1}{T}\right).
$$

Hence,

$$
\left\|\hat{\theta}_{r,r} - \theta_r(\beta_0)\right\|^2 = O_p\left(\frac{r}{T}\right),
$$

which is what we needed to show in Lemma 7.

Proof of Theorem 4 One can notice that

$$\lambda(\beta) = \lambda(\beta_0) - 2(\widehat{\theta}_{r,r} - \theta_r(\beta_0))'Q(\theta_r(\beta_0) - \theta_r(\beta)) + (\theta_r(\beta_0) - \theta_r(\beta))'Q(\theta_r(\beta_0) - \theta_r(\beta)). \tag{2.7.1}$$

According to Assumption 2,

$$\lim_{q\to\infty} (\theta_r(\beta_0) - \theta_r(\beta))'Q(\theta_r(\beta_0) - \theta_r(\beta)) = f(\beta).$$

Using Lemma 7, we can see that

$$\lambda(\beta_0) = \|Q^{1/2}(\widehat{\theta}_{r,r} - \theta_r(\beta_0))\|^2 = \|Q\|_1 \cdot O_p(r/T).$$

Now, we derive bounds on the middle term on the left side of equation (2.7.1):

$$\begin{aligned}
&\left|(\widehat{\theta}_{r,r} - \theta_r(\beta_0))'Q((\theta_r(\beta_0) - \theta_r(\beta))\right| \\
&\leq \|Q^{1/2}(\widehat{\theta}_{r,r} - \theta_r(\beta_0))\|\|Q^{1/2}(\theta_r(\beta_0) - \theta_r(\beta))\| \\
&\leq \|Q^{1/2}(\theta_r(\beta_0) - \theta_r(\beta))\|\sqrt{\|Q\|_1 \cdot O_p(r/T)}.
\end{aligned}$$

Finally,

$$\|Q^{1/2}(\theta_r(\beta_0) - \theta_r(\beta))\| \leq \|Q\|_1^{1/2}\|\theta_r(\beta_0) - \theta_r(\beta)\|.$$

Given assumption 2 (iv), we have

$$\lim_{r\to\infty} \|\theta_r(\beta_0) - \theta_r(\beta)\| \leq \sum_{j=r}^{\infty}(\|\Theta_j(\beta_0)\| + \|\Theta_j(\beta)\|) < \infty.$$

The reasoning above implies that under Assumptions 1 and 2, the stochastic process $\lambda(\beta)$ converges to $f(\beta)$ uniformly as $T \to \infty$ and $r = r_T$ changes according to Assumption 1. Assumption 2 guarantees identification, and by standard arguments Theorem 4 holds.

2.7.1.2 Proof of Theorem 5

In this proof, we mainly concentrate on the bias of the impulse response coefficient estimates $bias_{j,r} = E\hat{\Theta}_{j,r} - \Theta_j(\beta_0)$. The bias of $\hat{\Theta}_{j,r}$ comes from two sources: 1) $\theta_{j,r}$ is a non-linear function of $\{A_{i,r}\}_{i=1}^{r}$; and 2) the OLS estimates $\hat{A}_{i,r}$ are biased. Let $\Theta_j = \Theta_{j,r} = f_{j,r}(A_{1,r}, ..., A_{r,r})$ and $\widehat{\Theta}_{j,r} = f_{j,r}(\widehat{A}_{1,r}, ..., \widehat{A}_{r,r})$, then

$$
\begin{aligned}
E(\widehat{\Theta}_{j,r} - \Theta_j) &= E(f_{j,r}(\widehat{A}_{1,r}, ..., \widehat{A}_{r,r}) - f(A_{1,r}, ..., A_{r,r})) \cong \\
&\cong \sum_{k=1}^{r} \frac{\partial f_{j,r}(A_{1,r}, ..., A_{r,r})}{\partial A_{k,r}} E(\hat{A}_{k,r} - A_{k,r}) + \\
&+ \sum_{m=1}^{r} \sum_{k=1}^{r} \frac{\partial^2 f(A_{1,r}, ..., A_{r,r})}{\partial A_{k,r} \partial A_{m,r}} E(\hat{A}_{k,r} - A_{k,r})(\hat{A}_{m,r} - A_m) \\
&= bias_{j,r}^{(1)} + bias_{j,r}^{(2)}.
\end{aligned}
$$

We note here that in this model, we are not faced with misspecification bias as the true process, AR(1), is a sub-class of the estimated process, AR(r). Also notice that the true coefficients are $A_{j,r} = A_j = \beta_0 \mathbb{I}\{j = 1\}$, and we drop the redundant index, r, here.

Both summands are of order $1/T$ as $T \to \infty$ and r and j stay constant. The question is how does the bias change as r increases, as well as if j is changing. The following three lemmas describe what happens to both bias terms and the variance term under the asymptotics defined in Assumption 3.

Lemma 8. *Let $0 < \beta_0 < 1 - \delta < 1$ and assumption 3 be satisfied, then there exist constants T_1, C_1, C_2, C_3, C_4 which may depend on β_0 but do not depend on r, j or T such that for all $T > T_1$:*

- *(1) $C_1/T \leq bias_{j,r}^{(2)} \leq C_2/T$ for $j \leq 2r$;*
- *(2) $C_3(j-2r)^2\beta_0^{j-2r}/T \leq bias_{j,r}^{(2)} \leq C_4(j-2r)^2\beta_0^{j-2r}/T$ for $j > 2r$.*

Lemma 9. *Let $0 < \beta_0 < 1 - \delta < 1$ and assumption 3 hold, then there exist constants T_1, C_1, C_2, C_3, C_4 which may depend on β_0 but do not depend on r, j or T such that for all $T > T_1$:*

- *(1) $C_1/T \leq \omega_{j,r} \leq C_2/T$ for $j \leq r$;*
- *(2) $C_3(j-r)^2\beta_0^{2(j-r)}/T \leq \omega_{j,r} \leq C_4(j-r)^2\beta_0^{2(j-r)}/T$ for $j > r$.*

Lemma 10. *Let $0 < \beta_0 < 1 - \delta < 1$ and assumption 3 hold, then there exist constants T_1, C which may depend on β_0 but do not depend on r, j or T such that for all $T > T_1$:*

- *(1) $\|bias_{j,r}^{(1)}\| \leq C/T$ for $j \leq r$;*
- *(2) $\|bias_{j,r}^{(1)}\| \leq Cj\beta^{j-r}/T$ for $j > r$.*

We start with few preliminary lemmas that will be used throughout the rest of the proofs.

Lemma 11. *For $0 < x < 1$ the following is true:*

- *(a) $\sum_{j=0}^{k}(j+1)x^j = \frac{1}{(1-x)^2} - \frac{(k+2)x^{k+1}}{1-x} - \frac{x^{k+2}}{(1-x)^2}$*
- *(b) $\sum_{j=0}^{k}(j+1)(j+2)x^j = \frac{2}{(1-x)^3} - \frac{(k+3)(k+2)x^{k+1}}{(1-x)} - 2\frac{(k+3)x^{k+2}}{(1-x)^2} - 2\frac{x^{k+3}}{(1-x)^3}$*

Proof of Lemma 11

- (a) $\sum_{j=0}^{k}(j+1)x^j = \frac{d}{dx}\left(\sum_{j=0}^{k} x^{j+1}\right) = \frac{d}{dx}\left(\frac{x - x^{k+2}}{1-x}\right)$.
- (b) $\sum_{j=0}^{k}(j+1)(j+2)x^j = \frac{d^2}{dx^2}\left(\sum_{j=0}^{k} x^{j+2}\right) = \frac{d^2}{dx^2}\left(\frac{x^2 - x^{k+3}}{1-x}\right)$.

Lemma 12. [19] *If r is fixed, standard asymptotics suggest that*

$$AsyE\left[\left(\hat{A}_{j,r}-A_j\right)\left(\hat{A}_{i,r}-A_i\right)'\right]=\begin{cases}\frac{1+\beta_0^2}{T}, & i=j>1;\\ \frac{1}{T}, & i=j=1;\\ -\frac{\beta_0}{T}, & i=j+1 \text{ or } i=j-1;\\ 0, & \text{otherwise.}\end{cases}$$

Lemma 13. *In the considered example,*

$$\frac{\partial f_{j,r}(A_1,...,A_r)}{\partial A_k}=(j-k+1)^+\beta_0^{j-k},$$

$$\frac{\partial^2 f_{j,r}(A_1,...,A_r)}{\partial A_k\partial A_m}=\beta_0^{j-k-m}(j-m-k+2)^+(j-m-k+1)^+,$$

where $(\cdot)^+$ stands for the positive part of the expression.

Proof of Lemma 8 Using Lemma 12, we have

$$\begin{aligned}Tbias_{j,r}^{(2)}=&\sum_{k=1,m=1}^{r}\frac{\partial^2\Theta_{j,r}}{\partial A_k\partial A_m}TE(\hat{A}_{k,r}-A_{k,r})(\hat{A}_{m,r}-A_m)=\\ =&\frac{\partial^2 f_{j,r}}{\partial A_1^2}+(1+\beta_0^2)\sum_{k=2}^{r}\frac{\partial^2 f_{j,r}}{\partial A_k^2}-2\beta_0\sum_{k=2}^{r}\left(\frac{\partial^2 f_{j,r}}{\partial A_k\partial A_{k-1}}\right).\end{aligned}$$

Now, we apply Lemma 13. For $j>2r$, we have

$$\begin{aligned}Tbias_{j,r}^{(2)}=&\beta_0^{j-2}j(j-1)+(1+\beta_0^2)\sum_{k=2}^{r}\beta_0^{j-2k}(j-2k+1)(j-2k+2)\\ &-2\beta_0\sum_{k=2}^{r}\beta_0^{j-2k+1}(j-2k+2)(j-2k+3).\end{aligned}$$

[19] Lemma 12 is taken from Hamilton (1994, p.120).

Via calculations similar to those in Lemma 11, we get

$$Tbias_{j,r} = (j-2r+2)(j-2r+1)\beta_0^{j-2r} + 2\frac{\beta_0^{j-2r+2} - \beta_0^j}{1-\beta_0^2},$$

which gives us (2) of Lemma 8.

Next, let $j < 2r$. We consider two cases: j is even and j is odd. Let's start with the former, $j = 2r_1$. In this case, $\frac{\partial^2 f_{j,r}}{\partial A_{r_1}^2} = 2$, and

$$\begin{aligned} Tbias_{j,r}^{(2)} =& \beta_0^{j-2} j(j-1) + (1+\beta_0^2)\sum_{k=2}^{r_1} \beta^{j-2k}(j-2k+1)(j-2k+2) \\ &- 2\beta_0 \sum_{k=2}^{r_1} \beta_0^{j-2k+1}(j-2k+2)(j-2k+3) = \\ &= (j-2r_1+2)(j-2r_1+1)\beta_0^{j-2r_1} + 2\frac{\beta_0^{j-2r_1+2} - \beta_0^j}{1-\beta_0^2} = 2 + 2\frac{\beta_0^2 - \beta_0^j}{1-\beta_0^2}. \end{aligned}$$

For the odd case, $j = 2r_1 + 1$, we have $\frac{\partial^2 f_{j,r}}{\partial A_{r_1} \partial A_{r_1+1}} = 2$, and

$$\begin{aligned} Tbias_{j,r}^{(2)} =& \beta_0^{j-2} j(j-1) + (1+\beta_0^2)\sum_{k=2}^{r_1} \beta_0^{j-2k}(j-2k+1)(j-2k+2) \\ &- 2\beta_0 \sum_{k=2}^{r_1+1} \beta_0^{j-2k+1}(j-2k+2)(j-2k+3) \\ =& 6\beta_0 + 2\frac{\beta_0^3 - \beta_0^j}{1-\beta_0^2} - 4\beta_0. \end{aligned}$$

In both cases, one may put bounds on $Tbias_{j,r}^{(2)}$, which do not depend on T, j, or r but potentially depend on β_0. The statement (1) of Lemma 8 follows.

Proof of Lemma 9

$$TVar(\hat{\Theta_{j,r}}) =^{asy} \left(\frac{\partial \Theta_{j,r}}{\partial A_1}\right)^2 + (1+\beta_0^2)\sum_{k=2}^{r}\left(\frac{\partial \Theta_{j,r}}{\partial A_k}\right)^2 - 2\beta_0 \sum_{k=2}^{r}\left(\frac{\partial \Theta_{j,r}}{\partial A_k}\right)\left(\frac{\partial \Theta_{j,r}}{\partial A_{k-1}}\right)$$

For $j \leq r$, we have

$$TVar(\hat{\Theta_{j,r}}) =^{asy} j^2\beta_0^{2(j-1)} + (1+\beta_0^2)\sum_{k=0}^{j-2}(k+1)^2\beta_0^{2k} - 2\beta_0\sum_{k=0}^{j-2}(k+1)(k+2)\beta_0^{2k+1}$$
$$\to (1-\beta_0^2)\frac{2}{(1-\beta_0^2)^3} - (1+\beta_0^2)\frac{1}{(1-\beta_0^2)^2} = \frac{1}{1-\beta_0^2} \text{ as } j \to \infty,\ j \leq r.$$

The limit is finite and positive (bounded away from zero), and the convergence is uniform for $0 < \beta < 1 - \delta$. As a result, the right side of (1) from Lemma 9 holds. Lemma 9 (2) follows from the same type of arithmetic derivations as before.

Proof of Lemma 10 As the first step in this proof, we show that asymptotically

$$\sup_{1 \leq j \leq q_T} T|E\hat{A}_{j,r} - A_j| \leq C < \infty.$$

For this case, we extend Nicholls and Pope (1988), which computes the bias for a fixed r, to the case of r_T increasing. Following the proof in Nicholls and Pope (1988),[20] we get that the bias of the OLS estimates of equation (2.4.1) is

$$E(\hat{A}_{1,r}, ..., \hat{A}_{r,r})' - (\beta_0, 0, ..., 0)' = \tag{2.7.2}$$
$$= \frac{1}{T}\left(A\Gamma(0)B - \Gamma(0)\right)\left((I-B)^{-1} + B(I-B^2)^{-1} + \beta B(I-\beta B)^{-1}\right)\Gamma(0)^{-1}e_1 + O(T^{-3/2}), \tag{2.7.3}$$

where $A, B, \Gamma(0)$ are matrices of size $r \times r$, and $r = r_T$, while e_1 is an $r \times 1$ vector consisting of ones. Here,

$$A = \begin{pmatrix} \beta & 0 & \dots & 0 & \vdots & 0 \\ & & I_{r-1} & & \vdots & 0 \end{pmatrix}$$

is the matrix of the companion form $B = A'$, and $\Gamma(0)_{i,j} = \beta^{|i-j|}$. We used that the eigenvalues of B are $\{\beta, 0, ..., 0\}$, and the matrix which was called in Nicholls and Pope

[20]See equation (2.29) in Nicholls and Pope (1988).

(1988) $S(B)$ has the following form $S(B) = \beta(I - \beta B)^{-1}$.

Notice that the matrix $A\Gamma(0)B - \Gamma(0)$ has only one non-zero element, which is in the upper-left corner and equal to $-(1 - \beta^2)$. Next,

$$I - B^2 = \begin{pmatrix} 1-\beta^2 & -\beta & -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & -1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & -1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \end{pmatrix}.$$

From this, we can compute $(I - B^2)^{-1}$, which has 1's on the main diagonal and all even diagonals above the main diagonal, except that the first line is a repeating sequence of $\frac{1}{1-\beta^2}$ and $\frac{\beta}{1-\beta^2}$:

$$(I - B^2)^{-1} = \begin{pmatrix} \frac{1}{1-\beta^2} & \frac{\beta}{1-\beta^2} & \frac{1}{1-\beta^2} & \frac{\beta}{1-\beta^2} & \frac{1}{1-\beta^2} & \frac{\beta}{1-\beta^2} & \cdots \\ 0 & 1 & 0 & 1 & 0 & 1 & \ldots \\ 0 & 0 & 1 & 0 & 1 & 0 & \ldots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Now, notice that the matrix $(I - \beta B)^{-1}$ is upper triangular with $\frac{1}{1-\beta^2}, 1, 1, ..., 1$ on the main diagonal, and each element above the main diagonal is β times the element directly to its left. That is,

$$(I - \beta B)^{-1} = \begin{pmatrix} \frac{1}{1-\beta^2} & \frac{\beta}{1-\beta^2} & \frac{\beta^2}{1-\beta^2} & \frac{\beta^3}{1-\beta^2} & \cdots & \frac{\beta^{r-1}}{1-\beta^2} \\ 0 & 1 & \beta & \beta^2 & \ldots & \beta^{r-2} \\ 0 & 0 & 1 & \beta & \ldots & \beta^{r-3} \\ \vdots & \ddots & \ddots & \ldots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Matrix $(I - B)^{-1}$ is also upper triangular with all non-zero elements being 1, except that all elements in the first row are $\frac{1}{1-\beta}$:

$$(I-B)^{-1} = \begin{pmatrix} \frac{1}{1-\beta} & \frac{1}{1-\beta} & \frac{1}{1-\beta} & \frac{1}{1-\beta} & \cdots & \frac{1}{1-\beta} \\ 0 & 1 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & 1 & \dots & 1 \\ \vdots & \ddots & \ddots & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

In summary, matrix $(A\Gamma(0)B - \Gamma(0))\left((I-B)^{-1} + B(I-B^2)^{-1} + \beta B(I-\beta B)^{-1}\right)$ contains all zeros in all rows, except the very first one. Element j of the first row is given by $-(1+\beta) - \delta_j - \beta^j$, where $\delta_j = 1$ or $\delta_j = \beta$ for even and odd j respectively.

Finally, multiplying this by $\Gamma(0)$ we see that expression (2.7.2) contains all zeros in all rows, except the first, which is $-(1+\beta, 2-\beta, 1-2\beta, 2-\beta, 1-2\beta, 2-\beta, \beta, 1-2\beta, \ldots)$. Thus, for all β, there exists a constant C such that for all large T, and for all $j \leq r_T$, we have $T\|E(\hat{A}_{j,r} - A_{j,r})\| \leq C$.

Finally, recall that $bias^{(1)}_{j,r}$ is given by

$$bias^{(1)}_{j,r} = \sum_{k=1}^{r} \frac{\partial \Theta_{j,r}(A_1, \ldots, A_r)}{\partial A_k} E(\hat{A}_{k,r} - A_k).$$

The preceding results imply that $T\|bias^{(1)}_{j,r}\| \leq C \sum_{k=1}^{r} (j-k+1)^{+} \beta^{j-k}$.

Proof of Theorem 5 All statements below are for large enough T. We compare asymptotic bias with asymptotic variance of the expression in formula (2.4.3):

$$\sum_{j=1}^{q} \frac{E\left(\hat{\Theta}_{j,r} - \Theta_j\right)}{\omega_{j,r}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta} = \sum_{j=1}^{q} \frac{bias^{(1)}_{j,r} + bias^{(2)}_{j,r}}{\omega_{j,r}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta}.$$

First, consider only the bias due to the non-linearity of the impulse responses:

$$\sum_{j=1}^{q} \frac{bias_{j,r}^{(2)}}{\omega_{j,r}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta} \geq \sum_{j=2r}^{q} \frac{bias_{j,r}^{(2)}}{\omega_{j,r}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta}$$
$$\geq C \sum_{j=2r}^{q} \frac{\beta_0^{j-2r}(j-2r)^2/T}{\beta^{2(j-r)}(j-r)^2/T} j\beta_0^j = C \sum_{j=2r}^{q} \frac{(j-2r)^2}{(j-r)^2} j \geq$$
$$\geq C \frac{(q-2r)^2}{(q-r)^2} \sum_{j=2r}^{q} j \geq C(q-2r)^2.$$

The first inequality comes from dropping positive summands, the second follows from (2) of Lemma 9 and (2) of Lemma 8. The last two inequalities are simple algebra plus the assumption that $q > r \to \infty$. Here and in what follows, C is a general constant (different in different places) that may depend on β_0 but does not depend on q, r, or T.

For the other bias term, applying Lemma 10 gives

$$\sum_{j=1}^{q} \frac{bias_{j,r}^{(1)}}{\omega_{j,r}} \frac{\partial \Theta_j(\beta_0)}{\partial \beta} \leq C.$$

Next, looking at the variance of the objective function, we have that

$$\sum_{j=1}^{q} \frac{1}{\omega_{j,r}} \left(\frac{\partial \Theta_j(\beta_0)}{\partial \beta} \right)^2 = \sum_{j=1}^{q} \frac{1}{\omega_{j,r}} j^2 \beta^{2j} \leq$$
$$\leq C \sum_{j=1}^{r} \frac{1}{1/T} j^2 \beta^{2j} + C \sum_{j=r+1}^{q} \frac{1}{(j-r)^2/T\beta^{2(j-r)}} j^2 \beta^{2j} \leq$$
$$\leq CT + CT \sum_{j=r+1}^{q} \frac{1}{(j-r)^2} j^2 \beta^{2r},$$

and

$$\sum_{j=r+1}^{q} \frac{j^2}{(j-r)^2} = \sum_{j=r+1}^{q} \frac{(j-r)^2 + 2r(j-r) + r^2}{(j-r)^2} \leq q + 2r\log(q-r) + Cr^2.$$

Assumption 3 implies that

$$\sum_{j=r+1}^{q} \frac{1}{(j-r)^2} j^2 \beta^{2r} \to 0,$$

and

$$\sum_{j=1}^{q} \frac{1}{\omega_{j,r}} \left(\frac{\partial \Theta_j(\beta_0)}{\partial \beta} \right)^2 \leq CT.$$

Given Assumption 3, we can see that the ratio of the bias to the square root of the variance diverges. This effectively concludes the proof.

Figure A1: Impulse Response Functions

Panel A: Policy Shock

(ACEL Benchmark Parameter Estimates)

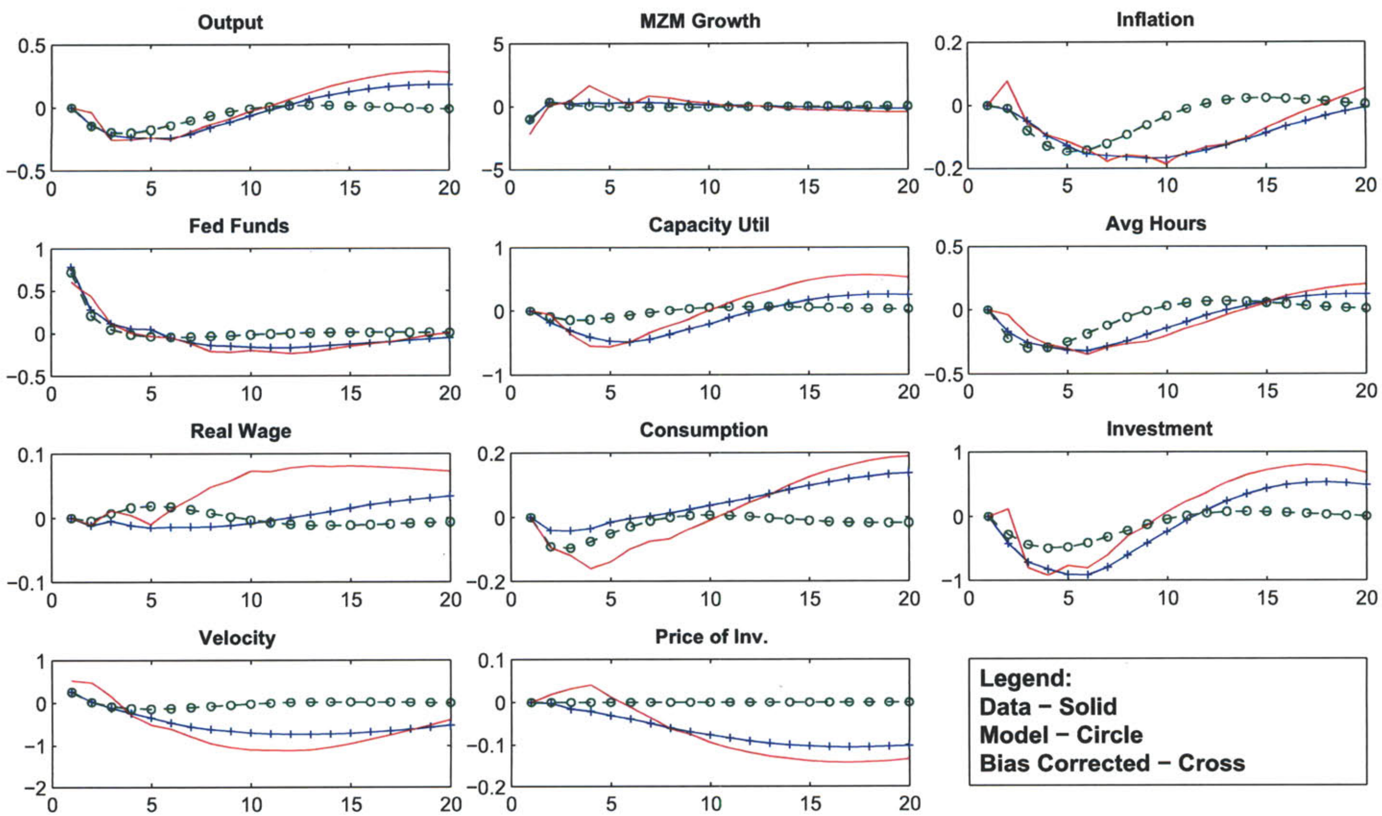


Figure A1: Impulse Response Functions

Panel B: Embodied Technology Shock

(ACEL Benchmark Parameter Estimates)

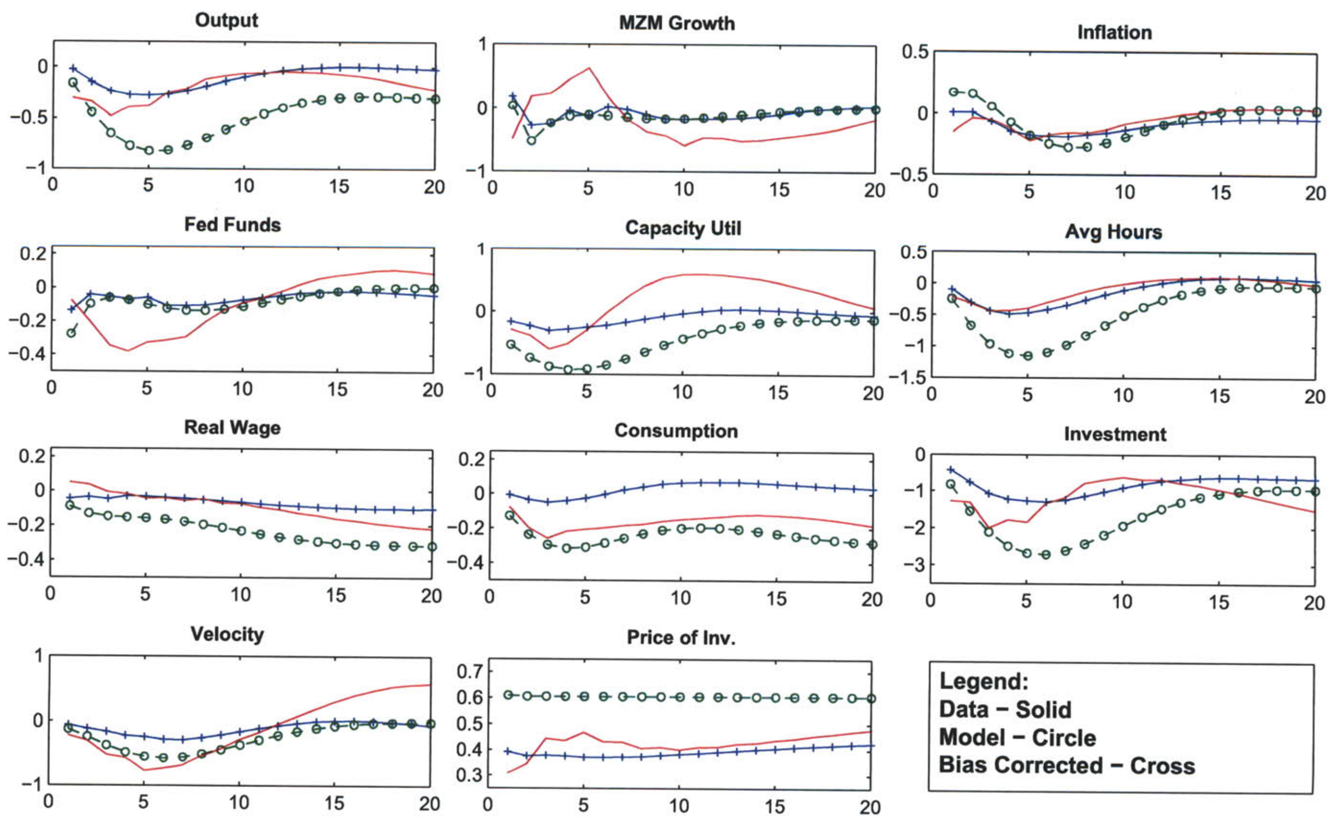


Figure A1: Impulse Response Functions

Panel C: Neutral Technology Shock

(ACEL Benchmark Parameter Estimates)

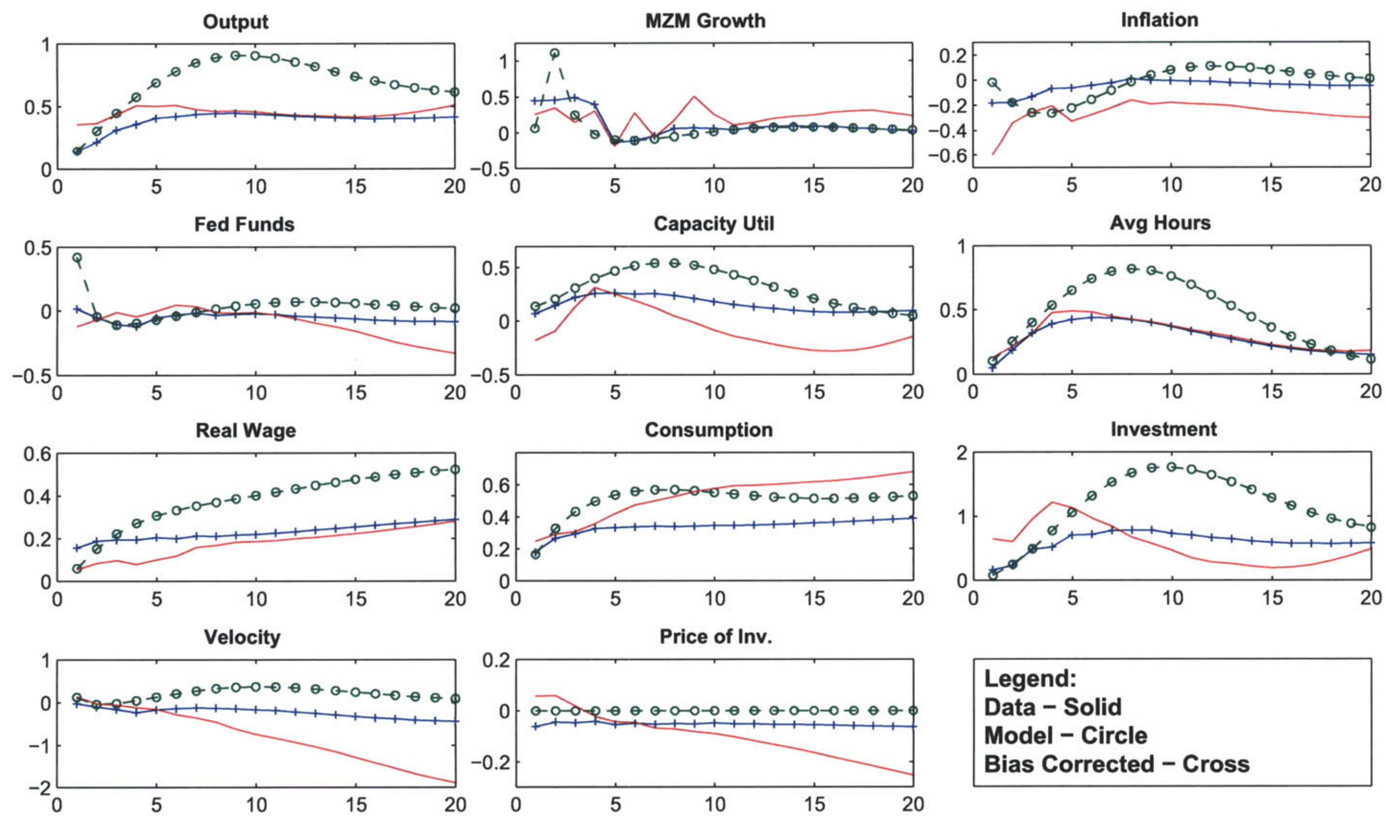


Table A1: ACEL Parameter Estimates

Approach:	Common	Bias Corrected			
	Baseline	Baseline	$W = I$	$q = 15$	Alternate
γ	0.04	0.83	0.89	1.20	0.01
ρ_M	-0.03	0.07	0.02	0.01	-0.18
ρ_{xz}	0.33	0.35	0.10	-0.04	0.25
c_z	3.00	0.46	1.50	1.40	8.10
ρ_{μ_z}	0.90	0.77	0.50	0.63	0.92
$\rho_{x\Upsilon}$	0.82	0.58	0.62	0.62	0.65
c_Υ	0.24	0.13	0.30	0.28	0.30
ρ_{μ_Υ}	0.24	-0.005	-0.01	0.003	-0.002
σ_M	0.33	0.17	0.22	0.14	0.18
σ_{μ_z}	0.07	0.12	0.17	0.14	0.06
σ_{μ_Υ}	0.30	0.61	0.91	0.72	0.56
ε	0.81	0.34	0.28	0.29	0.35
S''	3.28	1.00	1.00	0.99	2.10
ξ_w	0.72	0.74	0.59	0.57	0.58
b	0.70	0.49	0.49	0.49	0.49
σ_a	2.02	0.34	0.11	0.001	30.7
c_z^p	1.33	1.90	0.10	-0.29	3.10
c_Υ^p	0.13	0.28	0.01	0.03	0.23

Notes: The "Alternate" parameters are not estimated, but they produce an objective value that is not significantly different from the baseline estimates.

Chapter 3

Carry Trade and Systemic Risk: Why are FX Options so Cheap?[1]

3.1 Introduction

The high return of forex carry trade —i.e., investing in high interest rate currencies and funding it with low interest currencies— has led to an extensive literature documenting the "puzzle" and its robustness to a wide variety of controls.[2] This carry trade premium is not explained by traditional risk factors, such as those suggested by Fama and French (1993).[3] Moreover, several recent works have examined whether carry returns can be explained by crash risk and concluded that it cannot. Most prominently, Burnside et al. (2011) find that hedging the carry with ATM FX options leaves its returns unchanged, and therefore conclude that the crash risk exposure is not the source of the premium.[4]

The main contribution of this paper is to turn the puzzle on its head. We reconcile the

[1]This chapter is co-authored with Ricardo Caballero.

[2]The profitability of the carry trade strategy stems from the fact that high interest rate currencies tend to appreciate rather than depreciate, in constrast with the most basic implication of the uncovered interest parity condition. Academics have dubbed this phenomenon "the forward premium puzzle."

[3]See, e.g., Tables 2, 3 and 7 from Burnside et al. (2011).

[4]Farhi et al. (2009) uses currency options data to estimate that crash risk may account for roughly 25% of carry returns in developed countries, leaving plenty of the carry return unexplained.

past findings by showing that while the standard carry trade *is essentially compensation for systemic risk*, the corresponding bundle of *crash protection FX options are puzzlingly cheap.* In particular, we show that carry trade returns are highly correlated with the returns of a VIX rolldown strategy —i.e., the strategy of shorting VIX futures and rolling down its term structure— for individual currencies as well as for diversified portfolios.[5] We find that while typical carry trade strategies produce large returns, this is explained by its comovement with VIX rolldowns. On the other hand, portfolios of exchange rate options designed to hedge the carry provide a cheap form of systemic risk insurance. As a result, when the carry trade is hedged with exchange rate options, its return remains strongly significant even after controlling for its exposure to VIX rolldowns.

As a preview of our results, we run regressions of the form

$$\bar{z}_t = \alpha + x_t \beta + e_t$$

where $\bar{z}_t$ are the excess returns to a carry trade portfolio and x_t are the excess returns to VIX rolldowns.[6] Figure 1 plots the cumulative returns to carry strategies that place equal weights on each of the 25 countries in our sample against the required returns based on their exposure to VIX rolldowns.[7] Panel A shows the results for the standard carry, and Panel B is for the carry when hedged with at-the-money exchange rate options. Each panel includes two series: $\prod_{j=1}^{t}(1+\bar{z}_j)-1$ ("Realized") and $\prod_{j=1}^{t}\left(1+x_j\hat{\beta}_{j-1}\right)-1$ ("Required"). While the unhedged carry does only marginally better than its required return, the hedged carry beats its systemic counterpart by a wide margin. For unhedged carry, the realized Sharpe ratio is 0.43, only just barely higher than the 0.42 of its required returns. But for hedged

[5]The VIX is an S&P500 implied volatility index, which is often described as the "global (financial) fear" indicator.

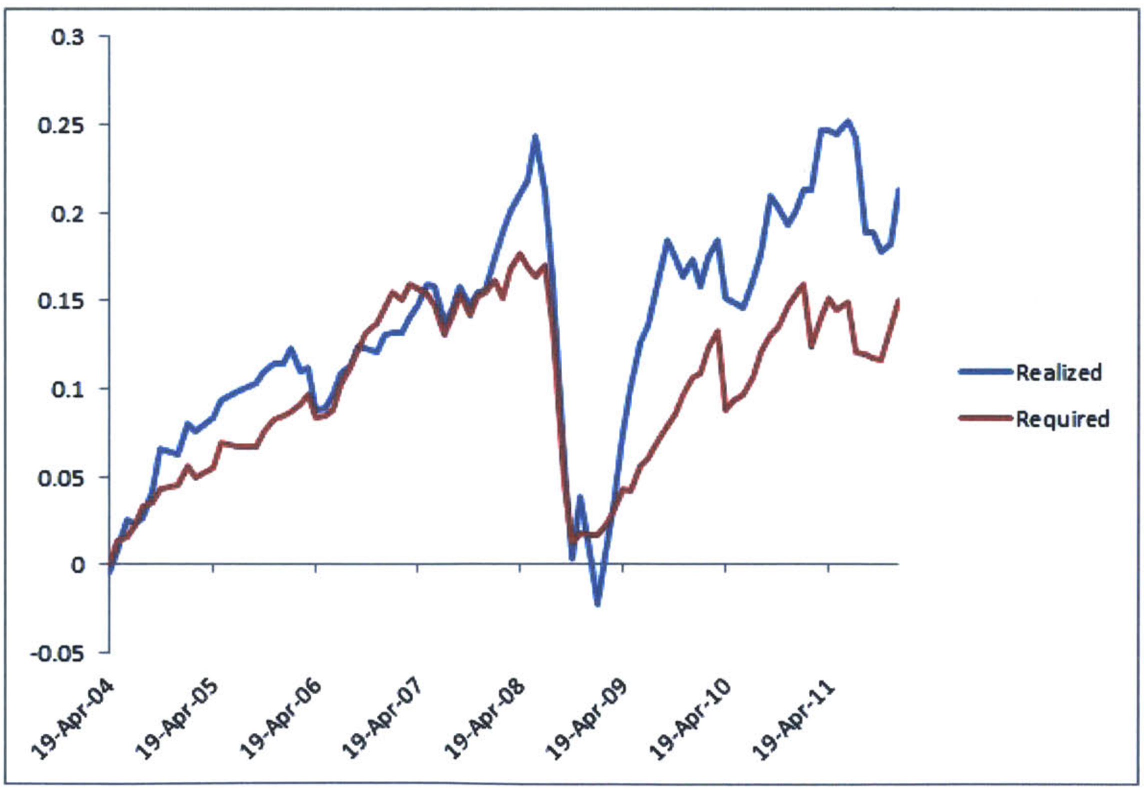
[6]Our main results use monthly data from March 2004 (the starting date for VIX futures) to January 2012.

[7]$\bar{z}_t = \frac{1}{N}\sum_{i=1}^{N} sign\left(r_{i,t} - r_{US,t}\right) z_{i,t}$ where $z_{i,t}$ are the (either hedged or unhedged) excess returns to the carry for currency i using USD as the base currency.

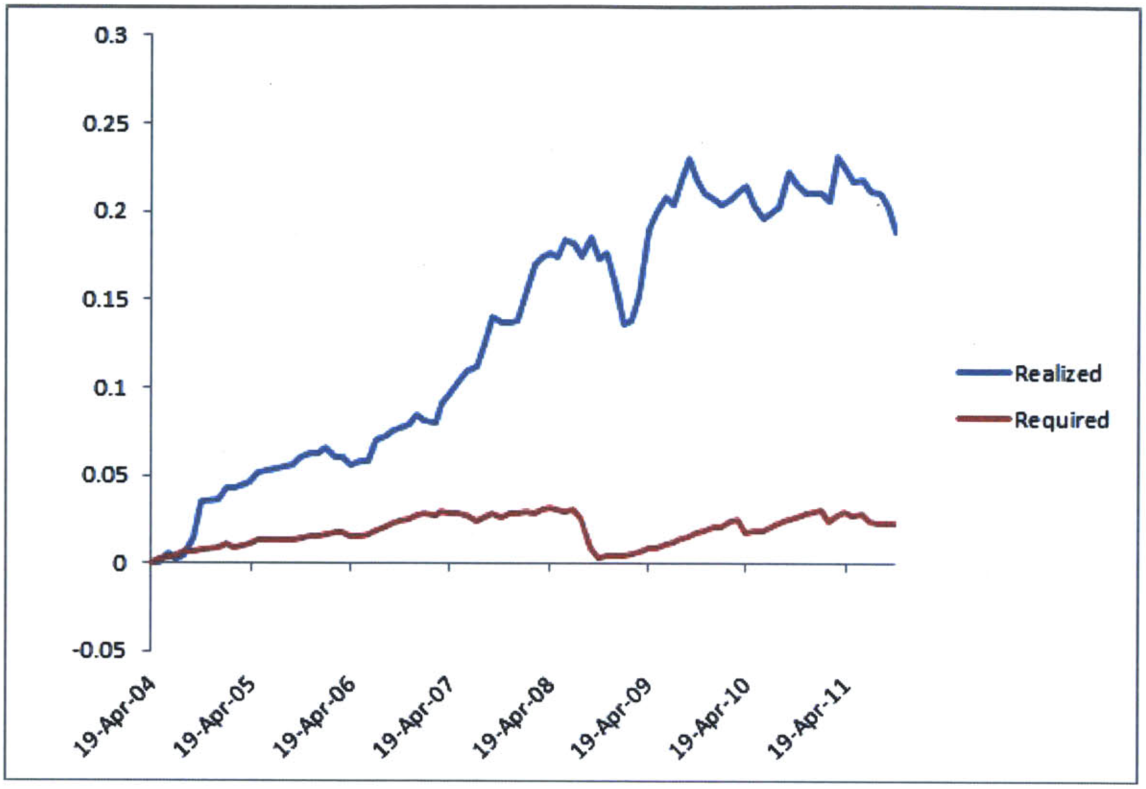
carry, the realized Sharpe ratio of 0.88 is more than double the 0.35 earned by its required. This suggests that exchange rate options earn a significant premium above that which would be required based on their systemic exposure. Indeed, we find that after controlling for its systemic exposure, the portfolio of carry protection options alone earns a Sharpe ratio of 1.14, and we thus conclude that such a portfolio of exchange rate options is a cheap form of systemic insurance.

Figure 1: Cumulative Returns

Panel A: Unhedged



Panel B: Hedged



Our work is directly related to a long literature that documents the puzzlingly high returns to the carry trade. This literature includes dozens, if not hundreds, of papers. Perhaps the most influential paper is Fama (1984) while Engel (1996) provides an excellent survey. Burnside et al. (2011) confirms that these findings hold in more recent data and with updated models and risk-factors. Focusing on the cross-section, Lustig and Verdelhan (2007) show that low interest rate currencies provide a hedge for domestic consumption growth risk, and this can explain why these currencies do not appreciate as much as the interest rate differential. However, Burnside (2007) shows that such a model does not have significant $\beta's$ and leaves a large unexplained intercept term, both of which are primary objects of our analysis.

Our findings also relate to a more recent literature that points to crash risk as a potential explanation for average carry returns. Jurek (2008) and Burnside et al. (2011) find that carry returns remain high even if one purchases crash protection via put options. Brunnermeier et al. (2009) shows that there is a strong relationship between carry return skewness and interest rate differentials, providing evidence that sudden exchange rate moves may be due to the reduction in carry positions as traders near constraints. Further, they find that increases in VIX coincide with reductions in carry positions, and a higher level of VIX predicts higher future carry returns. Focusing on the cross-section, Menkhoff et al. (2011) shows that there is a strong relationship between interest rates and global volatility measures. Finally, Lustig et al. (2010) shows that US business cycle indicators help predict foreign exchange returns.

The remainder of the paper is organized as follows. Section 2 introduces notation and discusses the basics of the carry trade. Section 3 describes the details of the theory that underlies our empirical approach. Section 4 provides a description of our data. Section 5 describes our results for the period leading up to the 2008 crisis and section 6 presents the results for the entire sample.

3.2 Carry Trade Basics

In this section, we describe standard carry trade strategies and introduce notation. For now, it is convenient to imagine that there are only two countries: the US (domestic) and foreign. Both countries issue sovereign risk-free bonds and have independent currencies. In the US, bonds pay an interest rate of r_t and abroad they pay r_t^*. Imagine also that there are only two periods, today and tomorrow. Let S_t be the exchange rate today and F_t be the forward exchange rate at which one can agree to make a currency exchange tomorrow. Both S_t and F_t are in terms of US dollars (USD) per foreign currency units (FCU). Without loss of generality assume that $r_t < r_t^*$. An obvious trading strategy for a Japanese investor today is to borrow \$1 at the low domestic interest rates, convert the borrowed USD to FCU, and lend those FCU to the foreign government at the higher interest rate. Since the investor borrows domestically, USD is referred to as the funding currency while FCU is the investment currency. Today the trader will have no net cash flows since he immediately lends all of his newly borrowed funds. Tomorrow the foreign government will pay him $\frac{(1+r_t^*)}{S_t}$ in FCU, and he will pay back $(1 + r_t)$ in USD. Since both interest rates are risk-free, the trader only faces risk from changes in the exchange rate. In order to perfectly hedge against FX risk, he could simply purchase a forward contract. Doing that, tomorrow he would receive the risk-free payment of $F_t \frac{(1+r_t^*)}{S_t}$ in USD. Since this strategy is both costless and riskless, any profits would be pure arbitrage. Therefore the following arbitrage condition is expected to be satisfied:

$$F_t \frac{(1+r_t^*)}{S_t} = 1 + r_t$$

This is known as the covered interest parity condition (CIP). In particular, it states that

$$f_t - s_t = \tilde{r}_t - \tilde{r}_t^*$$

where $f_t = log(F_t)$, $s_t = log(S_t)$, $\tilde{r}_t = log(1 + r_t)$, and $\tilde{r}_t^* = log(1 + r_t^*)$. The quantity $f_t - s_t$ is commonly referred to as the average forward discount, and we write $AFD_t = f_t - s_t$.

The carry trade is a simple risky variant on this strategy where the trader elects not to purchase the forward FX contract in the hope that the USD will not appreciate by enough to eliminate his entire profits. Denote the net payoff from this strategy by

$$\Pi_{t+1} = S_{t+1} \frac{(1+r_t^*)}{S_t} - (1 + r_t)$$

Using CIP, this can be rewritten as

$$\Pi_{t+1} = \frac{(1+r_t^*)}{S_t}(S_{t+1} - F_t)$$

That is, the carry trade strategy is equivalent to buying $\frac{(1+r_t^*)}{S_t}$ FCU forward. The trader will only lose money if the USD appreciates to the point that $S_t < F_t$, and if $E_t(S_{t+1}) > F_t$, the trader will make profits on average. In what follows, we will focus on the simpler and more commonly employed version of the carry trade where the investor buys 1 FCU forward. We define the excess return to this strategy as

$$z_{t+1} = \frac{S_{t+1}}{F_t} - 1.$$

3.3 Risk-based Explanations

As discussed in the introduction, empirical research has found the carry trade to produce puzzlingly high returns to investors. In principle, these returns could simply represent compensation for some form of risk, but it has proven difficult to identify which risks are relevant. In this section, we describe a simple and standard framework for assessing risk-based explanations, which we will rely upon for our analysis.

Standard models of asset prices can be reduced to the specification of an asset pricing kernel, M_{t+1}, such that

$$E_t[M_{t+1}R_{t+1}] = 0$$

holds for any excess return R_{t+1} denominated in USD. Plugging in $z_{t+1} = R_{t+1}$, we see that

$$\frac{E_t[M_{t+1}S_{t+1}]}{E_t[M_{t+1}]} = F_t$$

which implies

$$E_t[S_{t+1}] = F_t - \frac{Cov_t[M_{t+1}, S_{t+1}]}{E_t[M_{t+1}]}.$$

The historical literature, surveyed by Engle (1996), focuses on the special case where $Cov_t[M_{t+1}, S_{t+1}] = 0$. In that case, the forward rate is an unbiased predictor of the future spot rate, and the average excess carry return is zero. Both of these hypotheses have been consistently rejected empirically where researchers have estimated the regression

$$E_t[\frac{S_{t+1} - S_t}{S_t}] = \alpha + \beta\frac{F_t - S_t}{S_t}$$

and rejected the hypothesis that $\alpha = 0$ and $\beta = 1$. This is commonly referred to as the forward premium puzzle. In fact, estimates of β are often negative, implying that high interest rate currencies actually tend to appreciate, exactly the opposite from what many had expected.

Allowing for non-zero correlation between exchange rates and the pricing kernel opens the door for risk-based explanations. Rewriting things slightly, one can see that

$$E[Z_{t+1}] = \left(\frac{Cov[M_{t+1}, z_{t+1}]}{Var[M_{t+1}]}\right)\left(-\frac{Var[M_{t+1}]}{E[M_{t+1}]}\right)$$

or

$$E[z_{t+1}] = \tilde{\beta}\tilde{\lambda}$$

where $\tilde{\beta}$ is the slope coefficient from the regression of z_t on M_t and $\tilde{\lambda} = -\frac{Var[M_{t+1}]}{E[M_{t+1}]}$. Next if $M_{t+1} = a + bx_{t+1}$ where x_{t+1} is an excess return, then

$$\tilde{\lambda} = bE\left[x_{t+1}\right].$$

Therefore,

$$E[z_{t+1}] = \beta\lambda = \beta E\left[x_{t+1}\right]$$

where $\beta = b\tilde{\beta}$. In other words, the model predicts that if one runs the regression of $z_t = \alpha + \beta x_t + \epsilon_t$, one should find that $\alpha = 0$.[8] In other words, α is the portion of the excess return z_t that remains after controlling for its exposure to the factors x_t. The quantities $z_t - \hat{\beta}x_t$

[8] Tests of this form were first proposed by Black, Jensen, and Scholes (1972).

are commonly referred to as the pricing errors from the model where $\hat{\beta}$ is the coefficient estimated by OLS. The goal then becomes identifying a set of excess returns x_t that have this property. The classical example of this approach is Fama and French (1993) which shows that a set of three factors that do a very good job of pricing US equities. In this project, we consider using excess returns on VIX futures to price currency forwards.

3.4 Data

We obtained daily closing spot and forward exchange rates from Datastream for 67 countries in terms of USD/FCU. Data on VIX futures prices are also available from Datastream, and the VIX index can be downloaded from Yahoo! Finance. Our sample covers the period from March 26, 2004 when VIX futures started trading, to January, 2012. While FX forward rates trade for each horizon on every day, VIX futures only trade for fixed maturity dates. For each VIX futures contract, we find the trading day such that the maturity date of the contract is the same as the maturity date for the FX forward rates.[9] Since there is never more than one VIX futures expiration date in a single month, this procedure creates approximately non-overlapping holding periods at the one month horizon. In total, we have 89 observations for each country. We also employ daily data on the Fama-French three factors through November 30, 2011, which were obtained from Kenneth French's website. Daily returns are compounded to match the horizon under consideration. Our primary analysis focuses on a set of relatively developed countries, the Expanded Majors as defined by Bloomberg, and we also exclude the countries from this set which do not have floating exchange rates.[10] This leaves us with a total of 25 countries in our primary sample.[11]

[9] In cases when there is no exact match, we use the trading day such that the maturities are as close together as possible. They never differ by more than one day.

[10] We follow the IMF's classification of currencies, which is available at its website. We include only currencies which are classified by the IMF as either "Independently floating" or "Managed Floating with no pre-determined path for the exchange rate."

[11] Our primary sample of currencies includes AUD, BRL, CAD, CHF, CLP, COP, CZK, EUR, GBP, HUF, IDR, ILS, INR, KRW, MXN, NOK, NZD, PEN, PLN, SEK, SGD, TRY, TWD, USD, ZAR.

We also collected daily data on the implied volatilities of at-the-money (spot) exchange rate options with a one month horizon from Datastream for 22 out of the 25 countries in our primary sample.[12] We convert implied volatilities into prices using the Black-Scholes model. Our options data begins on March 26, 2004 and ends on November 30, 2011, leaving us with a total of 87 monthly observations.[13]

We consider two different periods of time for our analysis. Our primary period of interest includes only the dates in our sample prior to the financial crisis of 2008 (pre-crisis).[14] We focus on the pre-crisis period because most of the literature uses that period but also because in all likelihood the full sample dramatically overstates the probability of a catastrophic downturn in financial markets. However, we will show that the results from the pre-crisis period carry over qualitatively into the full sample period.

3.5 Carry Trade and Systemic Risk: Pre-Crisis

3.5.1 Traditional Carry Trade Portfolios

In this section, we begin by presenting results for the typical diversified carry trade strategies that have been the focus of much of the recent literature. These are portfolios of dynamically optimized carry strategies that are long a currency's forward whenever $r_t^* > r_t$ and short otherwise.[15] We follow Burnside et al. (2011) and others in using equally weighted baskets (EQL) with USD as the base currency. However, as pointed out by Jurek (2009), using interest rate spreads as portfolio weights (SPD) tends to produce higher returns, and the results may vary substantially for different base currencies. Therefore, we also consider

[12]The three excluded currencies are CZK, HUF, and PEN.

[13]Implied volatility data for TRY is missing for the first two observations in the sample.

[14]Specifically, we include monthly holding periods ending on May 19, 2004 through August 20, 2008 and only for those months when VIX futures contracts expired. This leaves a total of 48 observations.

[15]We assess whether $r_t^* > r_t$ by looking at a currency's average forward discount.

equal- and spread-weighted portfolios of JPY-based carry and currency neutral carry.[16] Our primary goal is to examine to what extent these returns can be explained by the excess returns to VIX futures. To that end, we run regressions of the form

$$\bar{z}_t = \alpha + x_t\beta + e_t$$

where $\bar{z}_t = \sum_{i=1}^{N} \omega_i sign\left(r^*_{i,t} - r_t\right) z_{i,t}$ is the excess returns to the carry trade portfolio, i indexes currencies, ω denotes portfolio weights, and x_t is the excess returns to a long position in VIX futures.

Table 1 summarizes the results from these regressions for the pre-crisis period. In every case, the excess carry trade returns were highly significant, consistent with the findings of Burnside et al. (2011). However, the exposure of these strategies to systemic risk is evident from the also highly significant values of β in all regressions. Importantly, after correcting for this risk exposure, we find little excess returns, as the estimated $\alpha's$ are statistically indistinguishable from 0 in all but one case.[17]

[16]Following Jurek (2009), currency neutral carry portfolios are computed in the following way. First, two sub-portfolios are formed one containing only those currencies with corresponding interest rates higher than that of the US and the other with only lower interest rate currencies. The final portfolio is equally weighted in these two sub-portfolios. The portfolio weighting scheme (EQL or SPD) refers to the weights used to construct the sub-portfolios.

[17]In the Appendix, we show that the results are robust to controlling for the Fama-French factors, and those traditional risk factors alone cannot explain the carry in line with the conclusions of Burnside et al. (2011).

Table 1: Carry Trade Returns and Systemic Risk (Pre-Crisis)

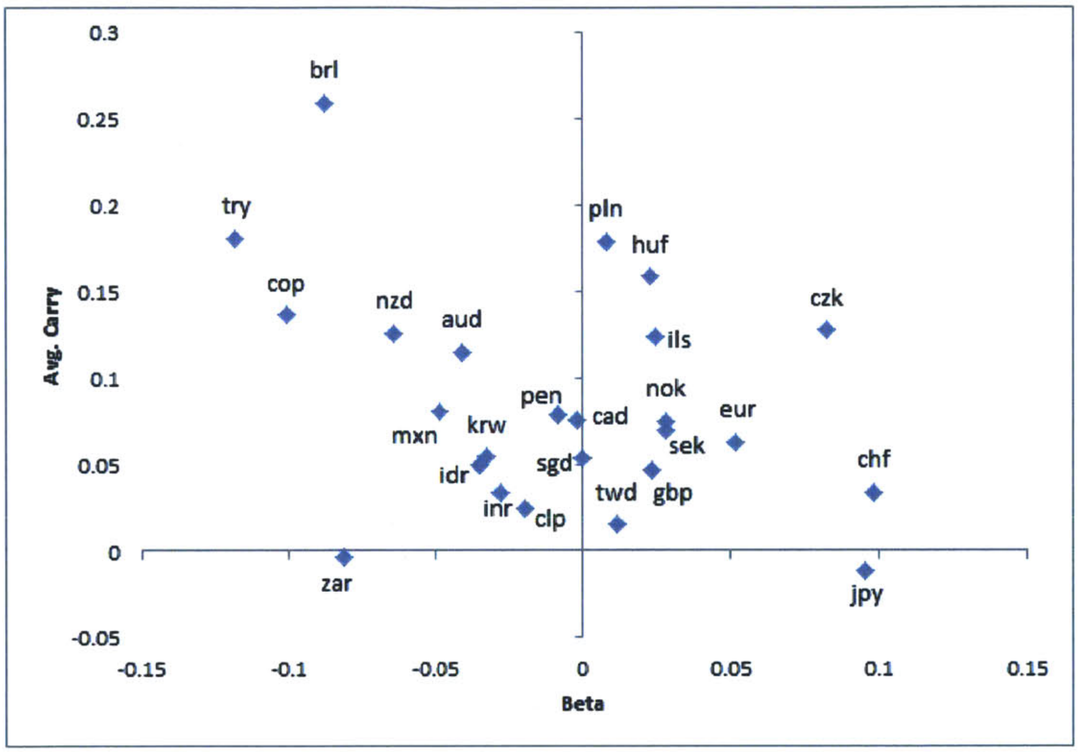
Currency	USD		JPY		Neutral	
Weights	EQL	SPD	EQL	SPD	EQL	SPD
Avg. Carry	4.86%***	12.07%***	10.63%**	15.42%***	3.79%***	7.92%***
	(2.85)	(4.51)	(2.31)	(2.82)	(2.76)	(3.62)
α	2.56%	7.86%**	2.70%	5.46%	0.94%	4.03%
	(1.28)	(2.62)	(0.56)	(0.95)	(0.61)	(1.58)
β	-0.028***	-0.052***	-0.097***	-0.122***	-0.035***	-0.048***

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - p<0.10, ** - p<0.05, *** - p<0.01.

These findings also carryover to individual currency pairs.

Figure 2: Avg. Carry vs. Beta (Pre-Crisis)

Panel A: USD



Panel B: JPY

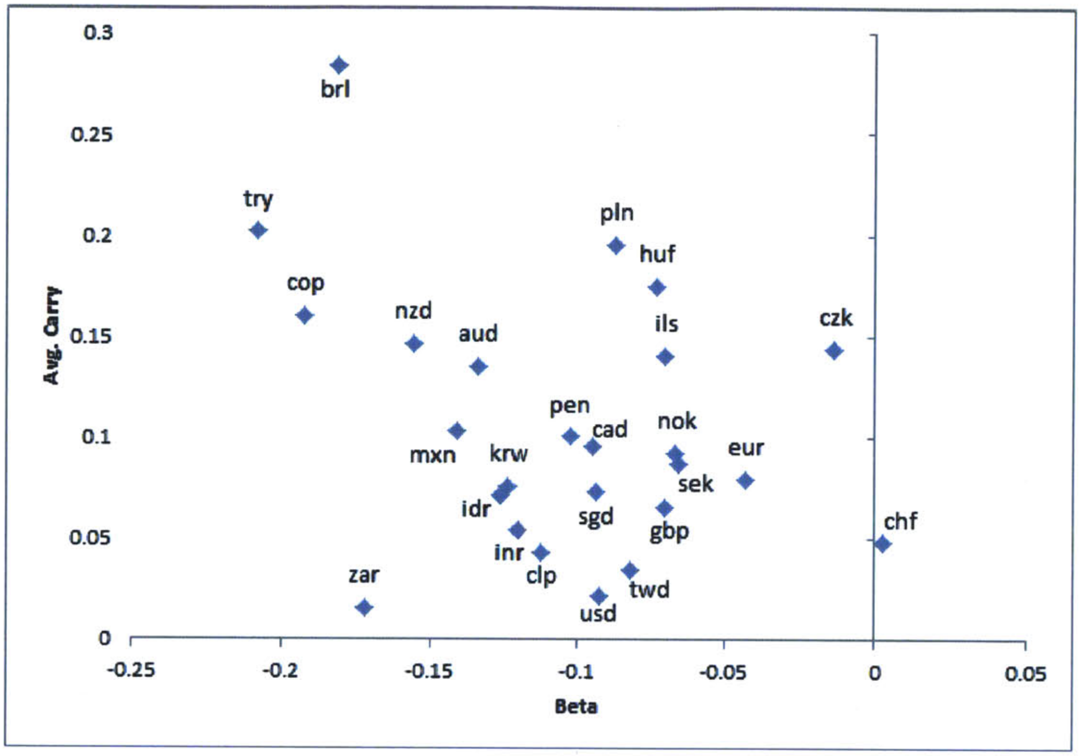
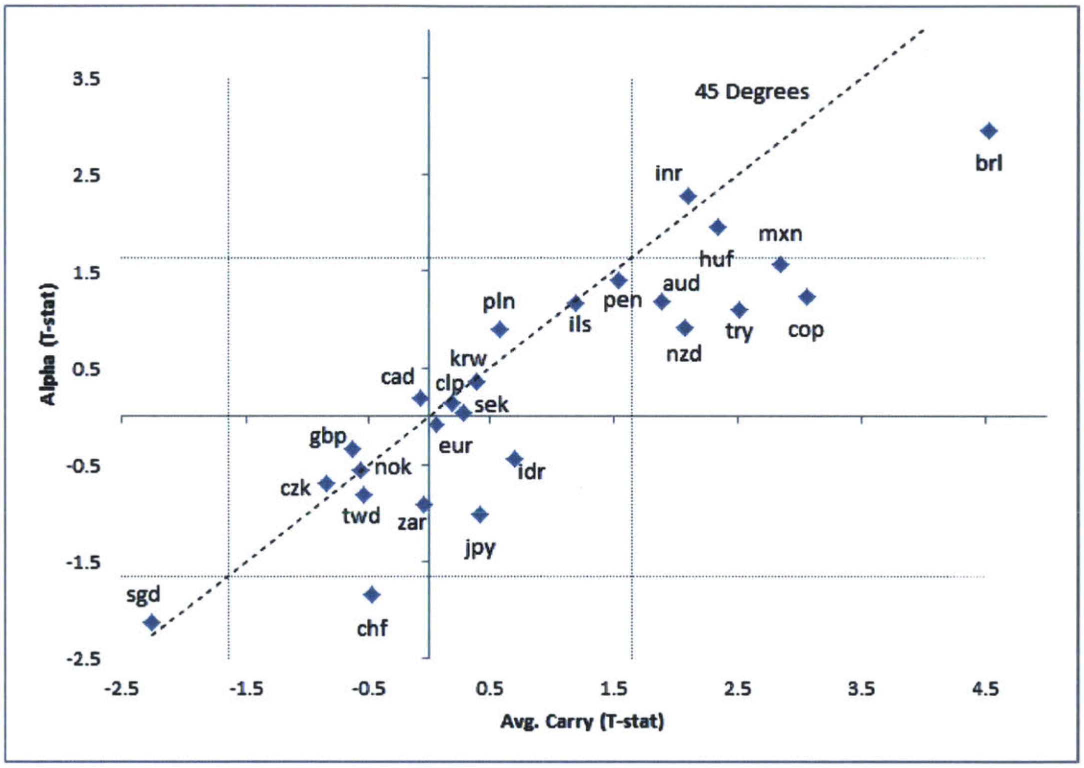


Figure 2 plots the average carry returns against the β_i's for positions that are all short the base currency (ie, the base currency is used as the funding currency). Note that this differs from the main strategy that we consider which determines the investment currency dynamically depending on the sign of the interest rate differential. Currencies which should have been funding currencies against the base show up here with positive β_i's. The evidence that carry is closely related to systemic risk exposure is visible.

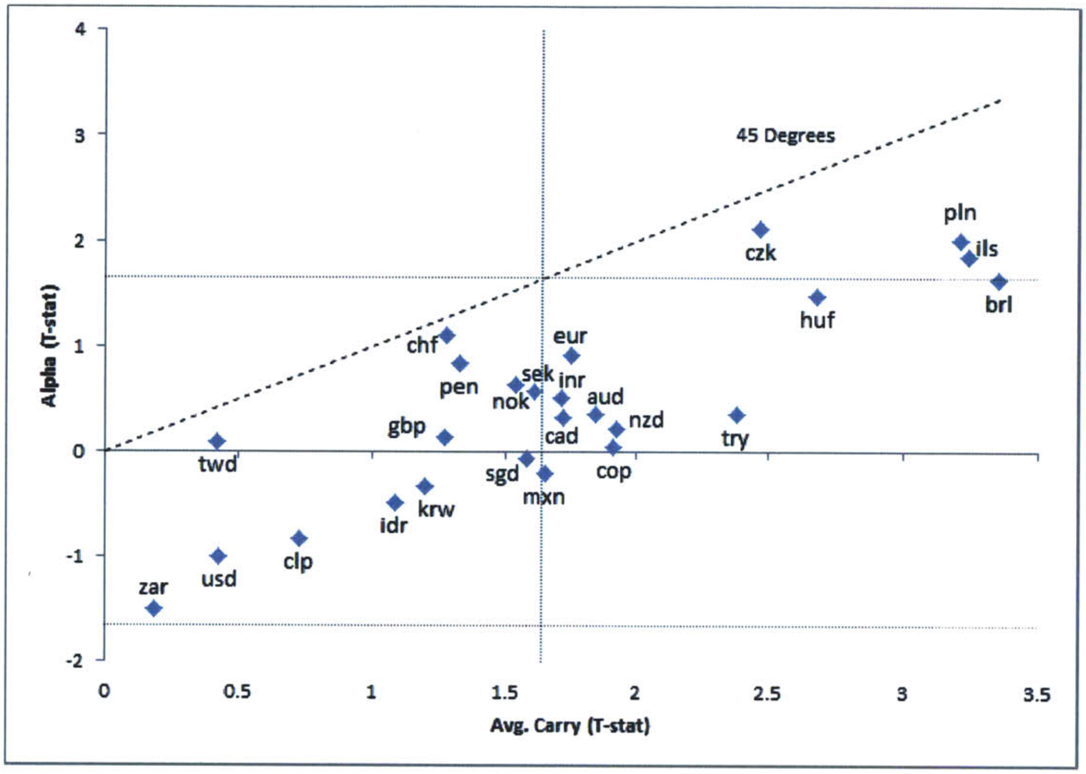
Figure 3: Alpha vs. Avg. Carry (Pre-Crisis)

Panel A: USD



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Panel B: JPY



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Figure 3 plots the T-statistics for average carry returns against the corresponding $\alpha's$ for each country in our sample. Here, funding currencies are determined dynamically by the sign of the interest rate differential. Results are shown separately when USD and JPY are used as the base currency. In line with our portfolio regressions, the $\alpha's$ are almost uniformly below the average carry returns across currencies. The pattern is most easily seen when JPY is used as the base where all currencies have positive carry returns with 13 significant at the 10% level, but only 3 $\alpha's$ are significantly above 0 while 7 are actually negative.

3.5.2 Options Hedged Carry Trade

Up to now, we have shown that hedging the carry trade with a long position in VIX rolldowns leaves investors with no significant excess returns. Next, we consider a natural alternative of hedging with exchange rate options. The strategy we examine involves purchasing an ATM call option on any exchange rate where the carry trader holds a short forward position and an ATM put on exchange rates where the carry trader is long. For 4 currencies in our sample (AUD, EUR, GBP, NZD), our options data is provided for the USD/FCU rate. In those cases, we compute the options hedged carry trade returns as[18]

$$z_t^h = \begin{cases} \frac{S_{t+1}-F_t+max\{0,S_t-S_{t+1}\}-(1+r)P_t}{F_t} & \text{if } S_t > F_t \\ \frac{F_t-S_{t+1}+max\{0,S_{t+1}-S_t\}-(1+r)C_t}{F_t} & \text{if } S_t < F_t \end{cases}.$$

In all other cases, the options data is for the FCU/USD rate, so we compute returns as

$$z_t^h = \begin{cases} \frac{S_{t+1}^{-1}-F_t^{-1}+max\left\{0,S_t^{-1}-S_{t+1}^{-1}\right\}-(1+r^*)P_t}{S_{t+1}^{-1}} & \text{if } S_t < F_t \\ \frac{F_t^{-1}-S_{t+1}^{-1}+max\left\{0,S_{t+1}^{-1}-S_t^{-1}\right\}-(1+r^*)C_t}{S_{t+1}^{-1}} & \text{if } S_t > F_t \end{cases}.$$

[19]

Our primary analysis involves the same regressions as before, but replacing z_t with z_t^h as the dependent variable. Table 2 displays the results. Consistent with the findings of Burnside et al (2011), hedging with FX options leaves a significant carry return. In fact, the T-statistics increase by over 70% on average. Furthermore, such hedging removes much of the trades' systemic exposure, reducing β by 64% on average and leaving significant $\alpha's$.

[18]This is the same formula used to compute FX options hedged carry by Burnside et al (2011).

[19]We compute this expression with S_{t+1}^{-1} in the denominator so that removing the options components leaves the same unhedged returns from before.

Table 2: Options Hedged Carry Trade Returns (Pre-Crisis)

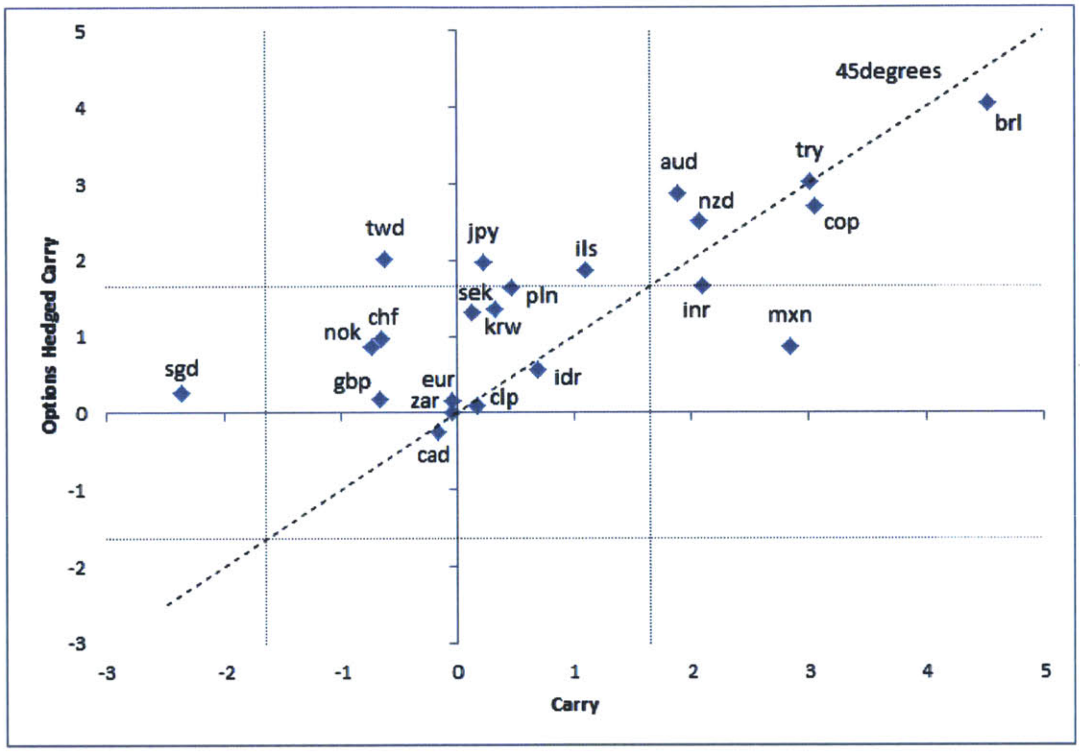
Currency	USD		Neutral	
Weights	EQL	SPD	EQL	SPD
Avg. Carry	4.47%**	10.69%***	3.38%**	6.77%***
	(2.61)	(3.81)	(2.30)	(3.04)
Hedged Carry	4.22%***	7.24%***	3.85%***	5.70%***
	(4.57)	(4.96)	(4.81)	(5.17)
α	3.41%***	5.12%***	2.74%***	3.85%***
	(3.49)	(4.24)	(3.79)	(4.39)
β	-0.01*	-0.02***	-0.01***	-0.02***

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - p<0.10, ** - p<0.05, *** - p<0.01.

Figure 4 shows that the results hold for individual currencies as well. The T-statistics for all options hedged returns are either close to or above those of their unhedged counterparts with the sole exception of MXN. The same is true in $\alpha - space$ where 18 out of 22 currencies' T-statistics are larger for the hedged strategy.

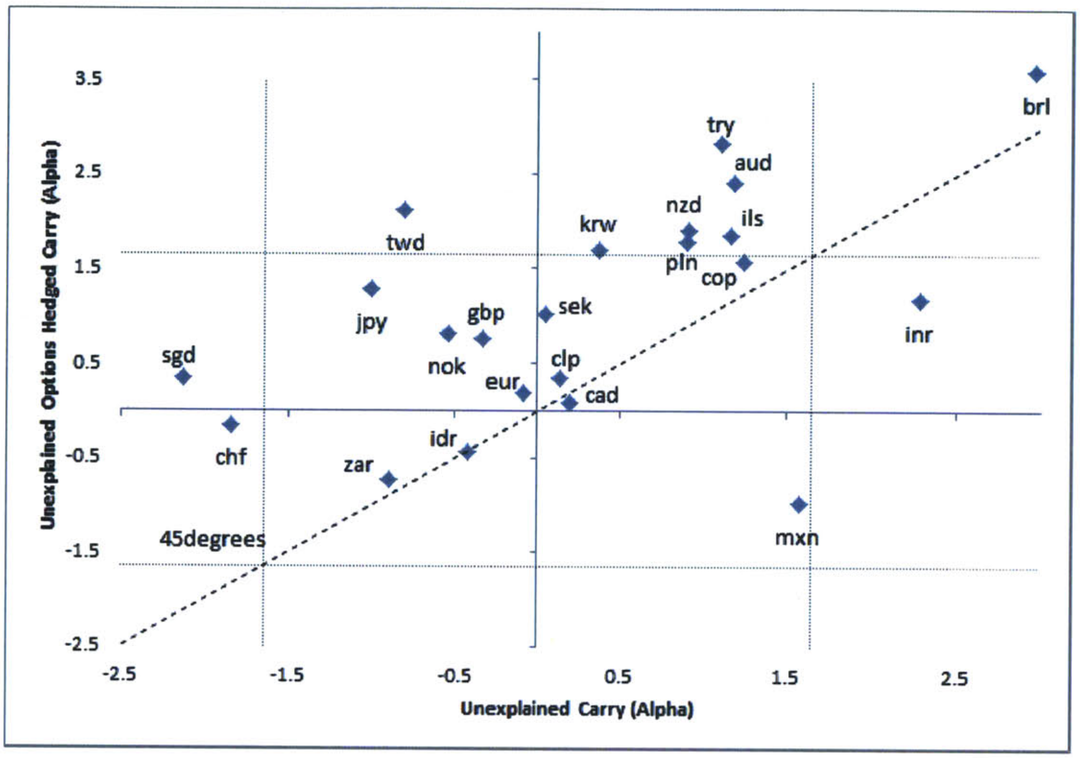
Figure 4: Hedged vs. Unhedged Carry (Pre-Crisis)

Panel A: Returns (T-stats)



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Panel B: Alphas (T-stats)

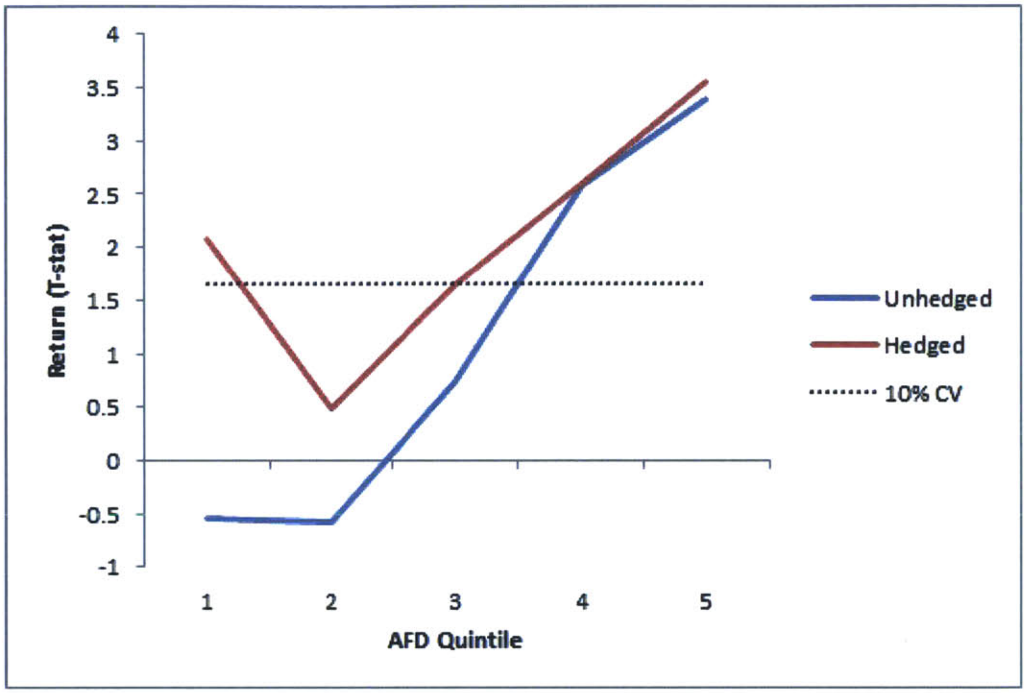


Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

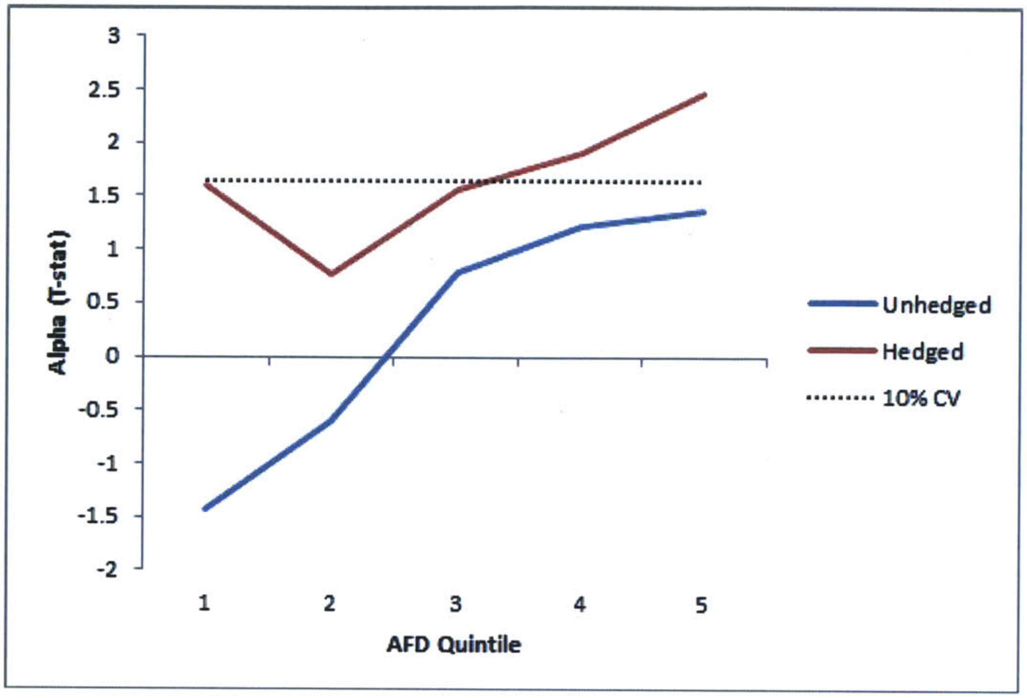
We dissect these results further by separating currencies into interest rate quintiles that are rebalanced every period. Figure 5 illustrates that the gains from hedging with options comes primarily from the low interest rate quintiles. For the two lowest quintiles, the carry trade returns were actually negative on average, but positive payoffs from options were enough to outweigh those losses. Meanwhile, at the two highest quintiles, Sharpe ratios were unchanged. In the $\alpha - space$, the gains still come primarily at the lowest quintiles where $\alpha's$ go from negative to positive. However, the highest quintiles also experienced some gains via a reduction in systemic risk exposure.

Figure 5: Carry Returns by Quintile (Pre-Crisis)

Panel A: Returns



Panel B: Alphas



Since the lowest quintile is comprised entirely of short positions in foreign currencies, we find that call options on funding currencies are especially cheap as both overall (Panel A) and systemic (Panel B) insurance.

3.5.3 A New Puzzle

Given our earlier results on the systemic exposure of unhedged carry returns, these fx-hedged findings suggest that a strategy involving only foreign exchange options designed to hedge carry returns may provide a cheap form of systemic insurance. We test this claim in the context of our model by regressing the excess returns to FX options on the excess returns to VIX rolldowns where the dependent variable is computed as

$$z_t^{opt} = \begin{cases} \frac{max\{0, S_t - S_{t+1}\} - P_t}{P_t} - r & \text{if } S_t > F_t \\ \frac{max\{0, S_{t+1} - S_t\} - C_t}{C_t} - r & \text{if } S_t < F_t \end{cases}$$

for those options quoted in terms of USD/FCU and

$$z_t^{opt} = \begin{cases} \left(\frac{S_{t+1}}{S_t}\right)\left(\frac{max\left\{0, S_t^{-1} - S_{t+1}^{-1}\right\} - P_t}{P_t}\right) - r & \text{if } S_t < F_t \\ \left(\frac{S_{t+1}}{S_t}\right)\left(\frac{max\left\{0, S_{t+1}^{-1} - S_t^{-1}\right\} - C_t}{C_t}\right) - r & \text{if } S_t > F_t \end{cases}$$

for those options quoted in FCU/USD. Estimates are contained in Table 3 for various portfolios of these options. In every case, the portfolio of options is significantly *positively* correlated with a long position in VIX rolldowns. This is not surprising as the options strategies are designed to hedge the carry, however the important result is that these portfolios have very large $\alpha's$ that are strongly significant in three out of four case, indicating that they do indeed provide (excessively) cheap systemic insurance.

Table 3: FX Options Returns and Systemic Risk (Pre-Crisis)

Currency	USD		Neutral	
Weights	EQL	SPD	EQL	SPD
α	467.75%***	287.30%	603.52%***	558.81%**
	(2.69)	(1.38)	(3.41)	(2.66)
β	2.14**	3.02***	2.53***	3.05***

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - p<0.10, ** - p<0.05, *** - p<0.01.

That is, the new puzzle is that there is premium to selling systemic risk insurance, hedged with currency options. The conventional carry trade is a form of selling systemic insurance, which when hedged with fx-options generates an excess return. But the source of the excess return is in the low cost of the hedge, not on the high return of the carry itself.

3.6 Carry Trade and Systemic Risk: Full Sample

Having examined the properties of typical carry trade strategies when hedged with either VIX futures or FX options during (relatively) normal times, it remains to be seen how these hedges perform in a sample that contains a major financial turndown. In this section, we explore this issue by using our full sample period that extends to January 2011 and includes the major asset market collapse surrounding the bankruptcy of Lehman Brothers in 2008. Although there are some nuances, the core message of the previous section remains unchanged.

Table 4 presents our factor regressions for the unhedged carry over this period. As noted earlier, our full sample overstates the empirical frequency of crises, and the result is that we do not see the strikingly positive carry trade returns as in the pre-crisis period or as has been documented for longer periods of time in previous work. On the other hand, as in the pre-crisis period, the exposure of the carry to systemic risk remains strongly significant in all cases, and taking this exposure into account, leaves the carry with broadly reduced returns and T-statistics. The JPY based carry even underperforms relative to its systemic exposure. Figure 6 shows that these results are true at the currency level as well where 21 out of 25 JPY based and 11 out of 25 USD based currency carry trades underperform.

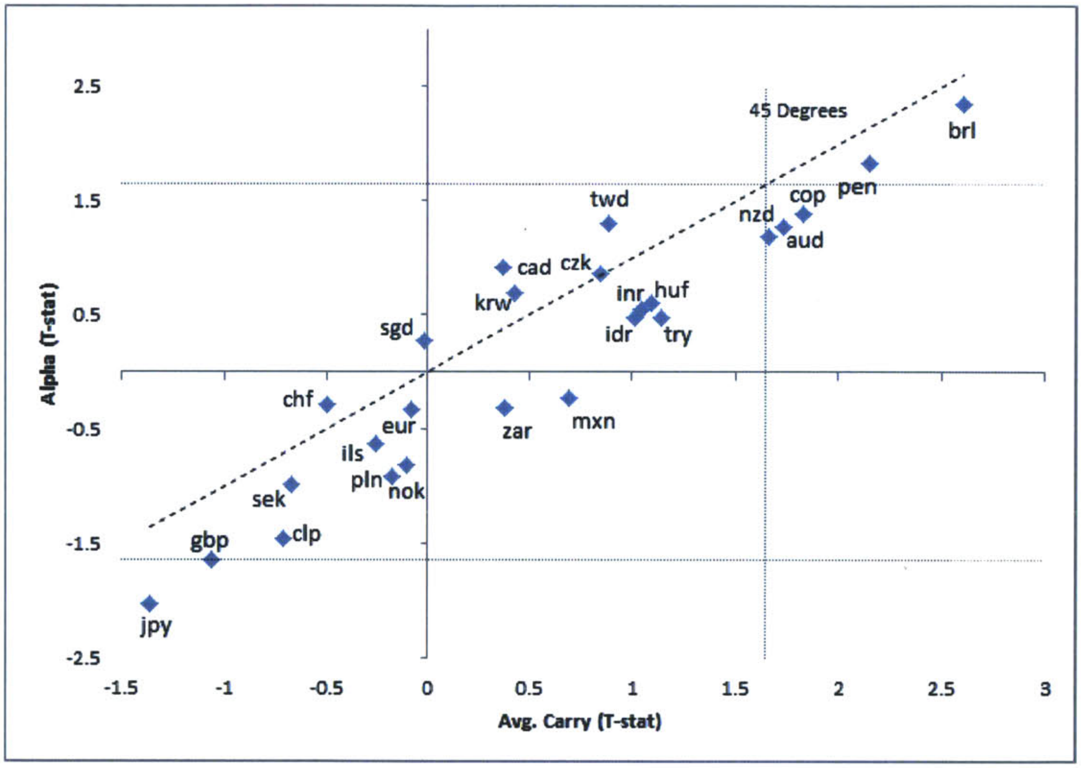
Table 4: Carry Trade Returns and Systemic Risk (Full Sample)

Currency	USD		JPY		Neutral	
Weights	EQL	SPD	EQL	SPD	EQL	SPD
Avg. Carry	2.82%	8.22%**	1.06%	4.04%	1.96%*	5.52%***
	(1.17)	(2.24)	(0.23)	(0.72)	(1.72)	(3.14)
α	0.83%	5.17%*	-2.91%	-0.86%	1.08%	4.15%***
	(0.47)	(1.98)	(-0.90)	(-0.23)	(1.21)	(2.98)
β	-0.049***	-0.074***	-0.097***	-0.120***	-0.021***	-0.033***

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - p<0.10, ** - p<0.05, *** - p<0.01.

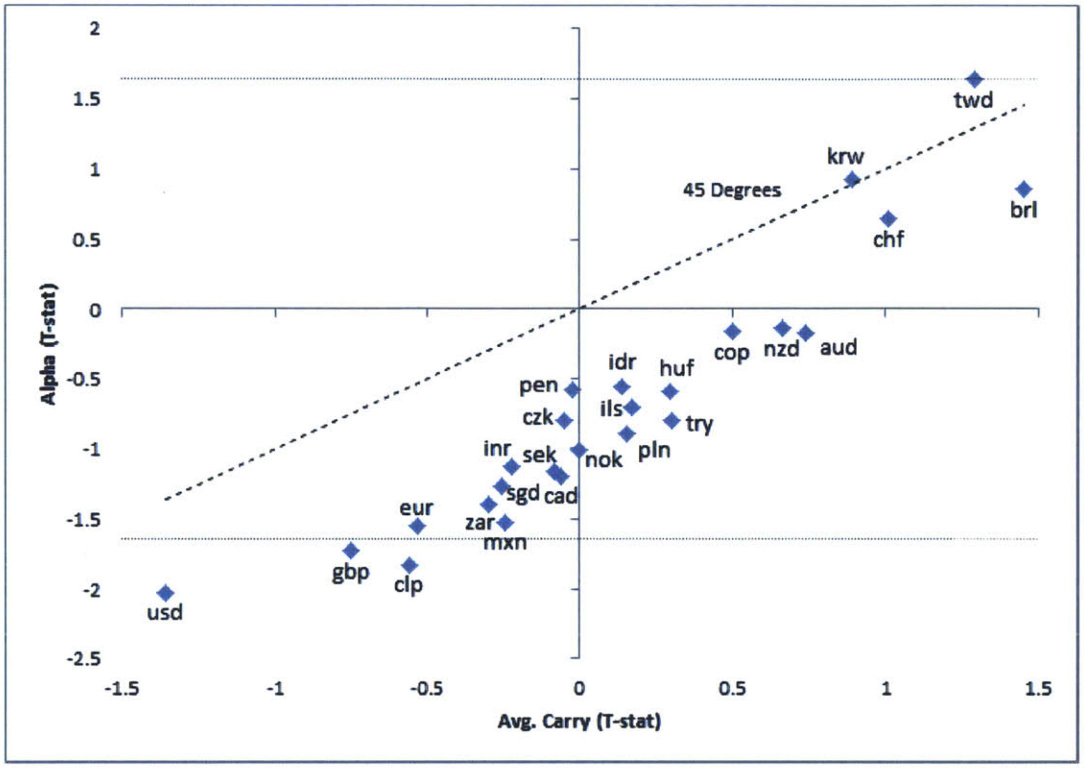
Figure 6: Alpha vs. Avg. Carry (Full Sample)

Panel A: USD



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Panel B: JPY



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Turning next to the options hedged carry, regression results are shown in Table 5. As before, the hedged carry is significant and remains so after correcting for its systemic exposure. This happens even though the unhedged carry is not broadly significant. The T-statistics for the hedged carry are on average 180% larger than those for the unhedged carry. Currency level regressions confirm these results where all but 4 currencies' T-statistics were higher for their hedged strategy, as illustrated in Figure 7.

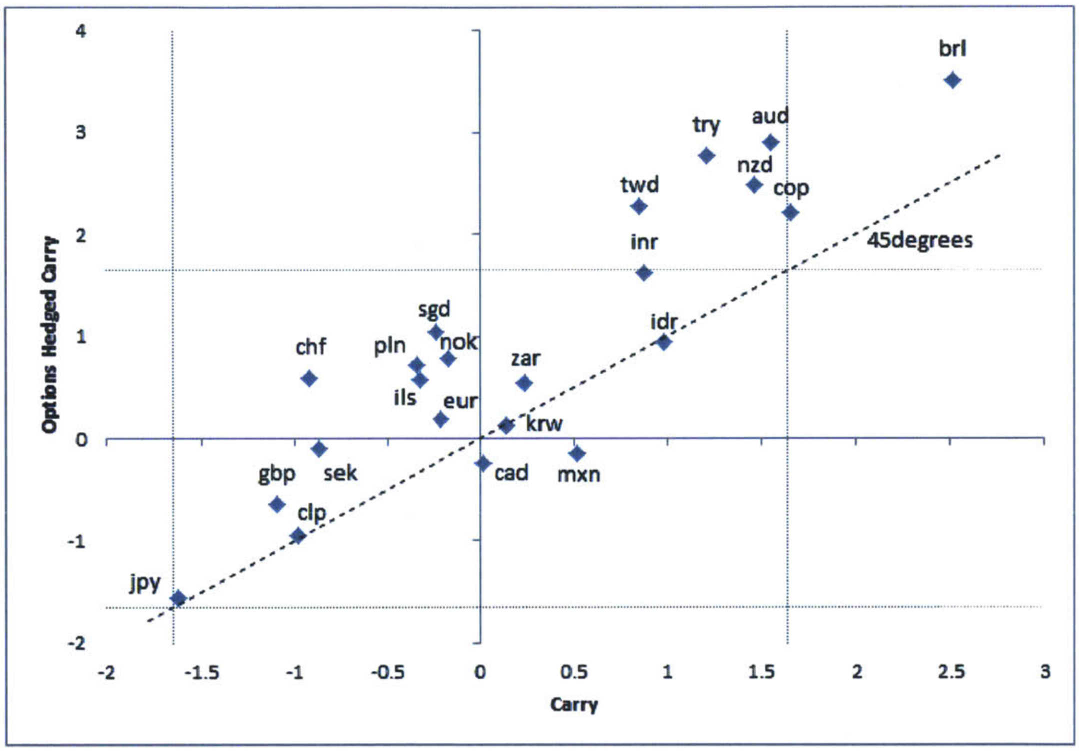
Table 5: Options Hedged Carry Trade Returns (Full Sample)

Currency	USD		Neutral	
Weights	EQL	SPD	EQL	SPD
Avg. Carry	1.58%	6.93%*	0.75%	4.04%**
	(0.64)	(1.87)	(0.56)	(2.15)
Hedged Carry	2.42%**	5.34%***	2.01%**	3.72%***
	(2.38)	(3.33)	(2.30)	(3.44)
α	1.77%*	3.78%**	2.20%**	3.55%***
	(1.69)	(2.47)	(2.01)	(2.79)
β	-0.01*	-0.02***	0.003	-0.002

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - $p<0.10$, ** - $p<0.05$, *** - $p<0.01$.

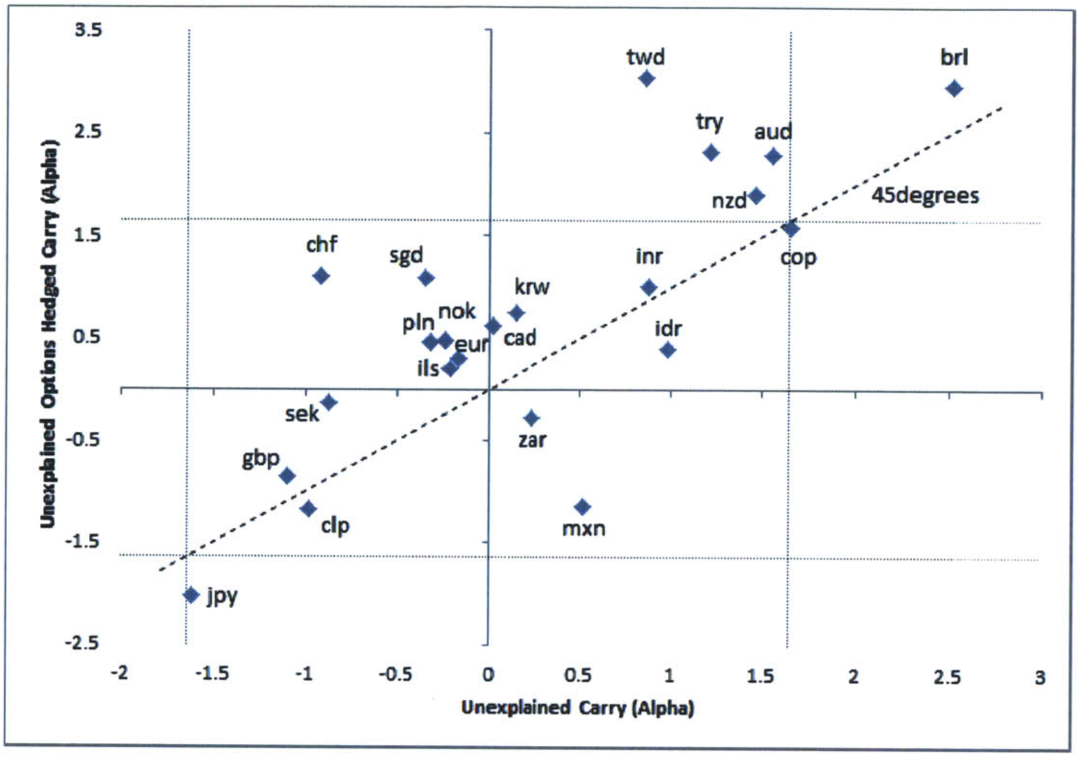
Figure 7: Hedged vs. Unhedged Carry (Full Sample)

Panel A: Returns (T-stats)



Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Panel B: Alphas (T-stats)

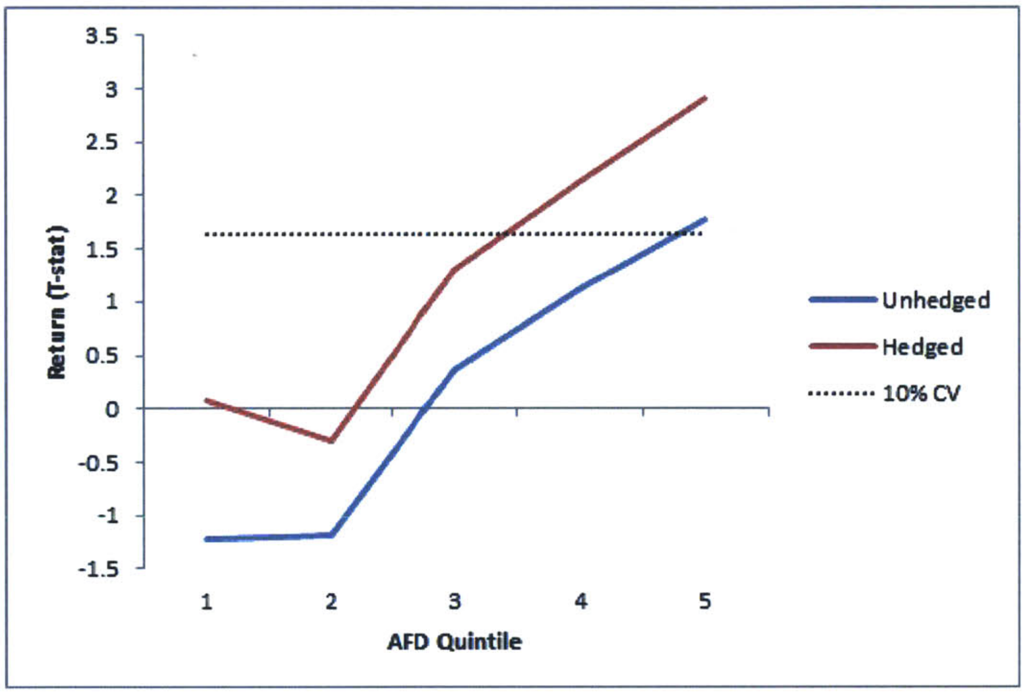


Notes: Dotted lines represent critical values for a two-sided test with a level of 10%.

Figure 8 plots the returns and $\alpha's$ separately for each interest rate quintile, illuminating a somewhat different pattern from pre-crisis. Whereas pre-crisis the gains from hedging were larger for lower interest rate quintiles, in the full sample they are spread quite evenly across interest rates. Panel B reveals that while all but one quintile of the unhedged carry actually underperformed its exposure, all but one quintile of the hedged carry outperformed. We also plot the cumulative returns to the two strategies in Figure 9. Interestingly, while the gains from hedging came largely from increased expected returns for lower quintiles, the gains for high quintiles were entirely from reduced volatility.

Figure 8: Carry Returns by Quintile (Full Sample)

Panel A: Returns



Panel B: Alphas

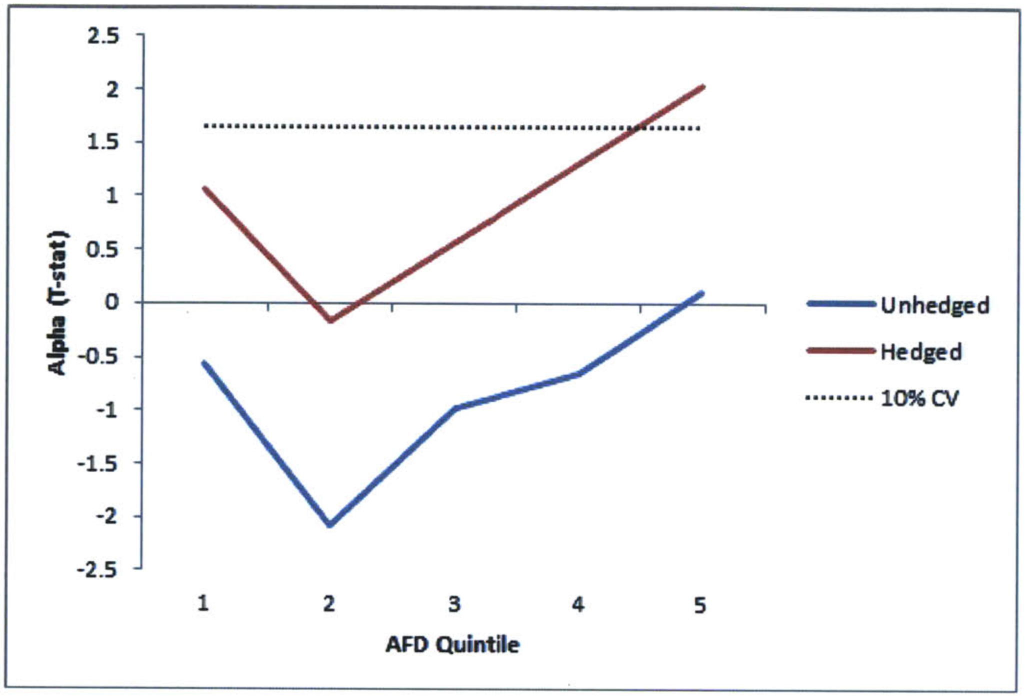
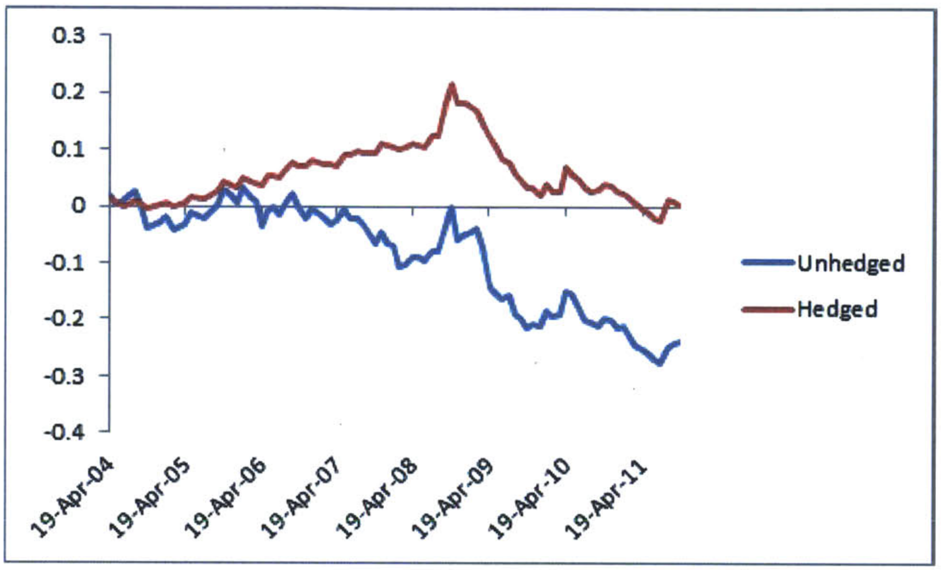
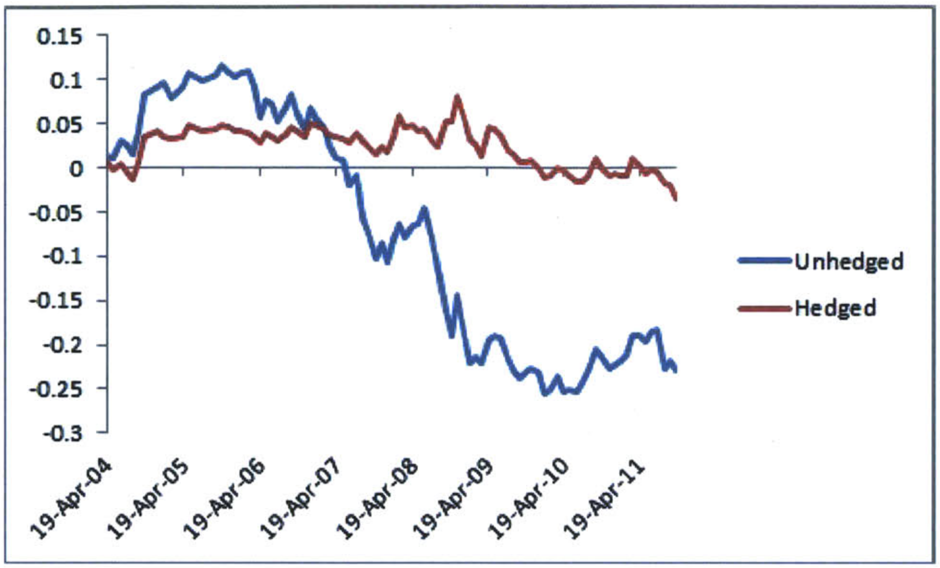


Figure 9: Cumulative Returns

AFD Quintile 1

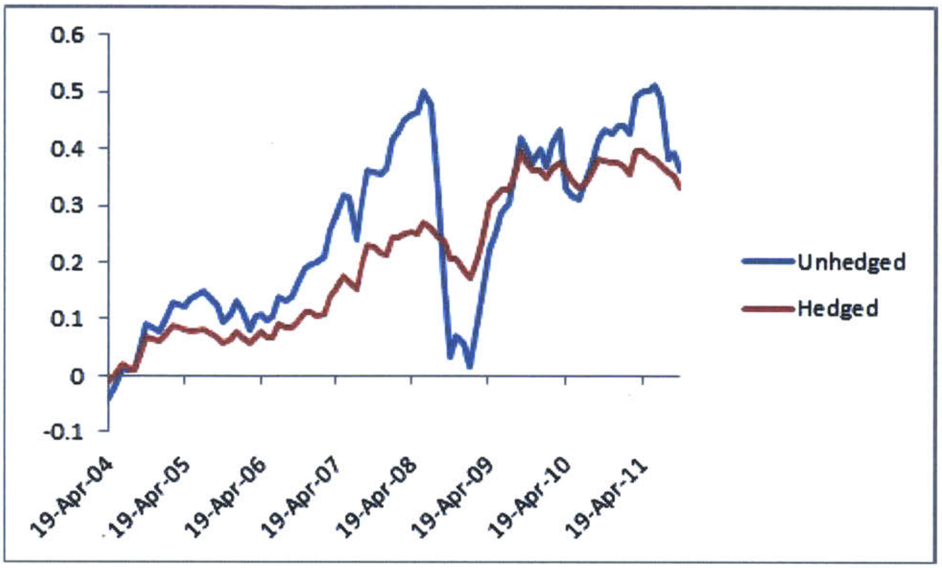


AFD Quintile 2

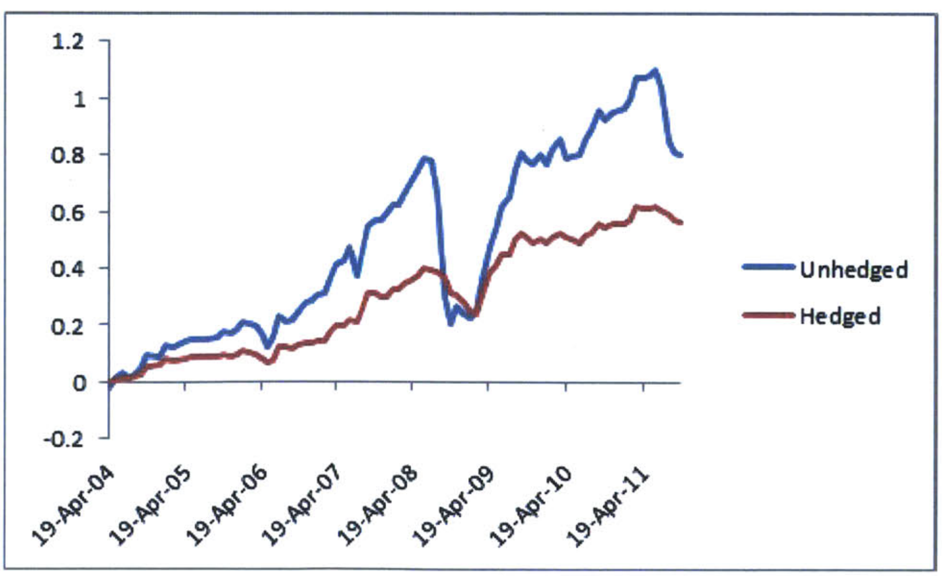


AFD Quintile 3

AFD Quintile 4



AFD Quintile 5



Finally, we implement our direct test of whether the appropriately constructed bundle of FX options provide a cheap form of systemic insurance. The results presented in Table 6 confirm our findings from the pre-crisis. In fact, the T-statistics for $\alpha's$ are even larger for the full sample.

Table 6: FX Options Returns and Systemic Risk (Full Sample)

Currency	USD		Neutral	
Weights	EQL	SPD	EQL	SPD
α	425.84%***	331.32%**	414.26%***	363.22%***
	(3.08)	(2.08)	(3.66)	(2.67)
β	3.13***	4.16***	2.13***	2.49***

Notes: Standard errors are robust to heteroskedasticity. Avg. carry and α are annualized. T-statistics are reported in parentheses. * - p<0.10, ** - p<0.05, *** - p<0.01.

3.7 Final Remarks

To summarize, we find that after appropriately hedging the carry trade with VIX roll-downs in both samples that exclude and include crises, there is no evidence that its return is particularly large, and in the latter, there is evidence that it may actually be too small. This contrasts with Burnside et al (2011) who find that the returns to the carry trade are not a compensation for risk. On the other hand, like those previous authors, we find that when the carry is hedged with FX options, it does indeed produce significantly positive returns for which we have no risk-based explanation. Taken jointly, our two sets of results suggest that portfolios of FX options designed to hedge the carry trade provide a relatively cheap means of hedging systemic risk, and our tests confirm this hypothesis.

Put differently, the new puzzle is that there is a premium to selling systemic risk insurance, hedged with currency options. The conventional carry trade is a form of selling systemic insurance, which when hedged with fx-options generates an excess return. But the source of the excess return is in the low cost of the hedge, not in the high return of the carry itself.

3.8 Appendix

3.8.1 Fama-French Risk Factors

In this section, we analyze the exposure of the carry to the traditional Fama-French risk factors (MKT, SMB, HML), as studied in Burnside et al. (2011). In the pre-crisis period, the estimates confirm that the carry trade is not significantly exposed to these traditional factors, and controlling for them leaves a significant excess return. Furthermore, we find that even after controlling for the Fama-French factors, the carry remains significantly exposed to VIX rolldowns, and the resulting α is still insignificant. In the full sample, the carry does have significant exposure to traditional risk factors, due to the heightened correlations between asset prices during the crisis. However, much of this exposure disappears once we control for VIX rolldowns.

3.8.1.1 Pre-Crisis

Table A1: Fama-French Risk Factors and Carry (USD, Pre-Crisis)

	(1)	(2)	(3)	(4)
MKT	0.075460	0.070145	-0.012829	-0.013387
SMB		0.050953		0.031169
HML		0.000064		-0.008894
β			-0.030732*	-0.029950*
α	0.003697**	0.003657**	0.002023	0.002068
	(2.49)	(2.58)	(1.20)	(1.22)
Adj.R2	0.061536	0.029792	0.129624	0.093736
N	48	48	48	48

Table A2: Fama-French Risk Factors and Carry (JPY, Pre-Crisis)

	(1)	(2)	(3)	(4)
MKT	0.324450***	0.331391***	0.102904	0.106104
SMB		-0.076004		-0.129363
HML		0.019656		-0.004504
β			-0.077117*	-0.080776*
α	0.007344**	0.007344*	0.003145	0.003059
	(2.06)	(1.99)	(0.73)	(0.67)
Adj.R2	0.190412	0.157283	0.249598	0.224398
N	48	48	48	48

Table A3: Fama-French Risk Factors and Carry (Neutral, Pre-Crisis)

	(1)	(2)	(3)	(4)
MKT	0.083448**	0.081659**	-0.039346	-0.038640
SMB		0.024933		-0.003560
HML		-0.016204		-0.029105
β			-0.042743***	-0.043133***
α	0.002768**	0.002797**	0.000440	0.000508
	(2.40)	(2.43)	(0.32)	(0.37)
Adj.R2	0.134863	0.101471	0.380201	0.356470
N	48	48	48	48

3.8.1.2 Full Sample

Table A4: Fama-French Risk Factors and Carry (USD, Full Sample)

	(1)	(2)	(3)	(4)
MKT	0.223386***	0.228638***	0.129237***	0.101508*
SMB		-0.052582		-0.025427
HML		0.012781		0.070356
β			-0.025779***	-0.029696***
α	0.001543	0.001566	0.001044	0.000942
	(1.06)	(1.08)	(0.75)	(0.67)
Adj.R2	0.515900	0.507569	0.560358	0.560638
N	87	87	87	87

Table A5: Fama-French Risk Factors and Carry (JPY, Full Sample)

	(1)	(2)	(3)	(4)
MKT	0.410837***	0.440020***	0.152276**	0.107427
SMB		-0.158055		-0.087013
HML		-0.015450		0.135176
β			-0.070796***	-0.077690***
α	-0.000360	-0.000259	-0.001731	-0.001892
	(-0.12)	(-0.09)	(-0.66)	(-0.71)
Adj.R2	0.467330	0.462354	0.562664	0.564331
N	87	87	87	87

Table A6: Fama-French Risk Factors and Carry (Neutral, Full Sample)

	(1)	(2)	(3)	(4)
MKT	0.083254***	0.089165***	0.018511	0.009950
SMB		0.010787		0.027707
HML		-0.030721		0.005154
β			-0.017727***	-0.018504***
α	0.001189	0.001200	0.000845	0.000811
	(1.51)	(1.54)	(1.12)	(1.06)
Adj.R2	0.328130	0.322564	0.430265	0.420669
N	87	87	87	87

Chapter 4

Reputation in Pay-as-You-Go Auctions

4.1 Introduction

The recent boom in online auction websites has produced a large amount of data on user behavior in auctions that has yet to be fully examined or explained. In this paper, I study a new auction format known as Pay As You Go Auctions (PAYGA, also known as penny auctions). In PAYGA, users pay for each bid that they place in an attempt to win the right to purchase an item at a price that is significantly below its market value. PAYGA has gained attention, as its popularity among users has grown rapidly in recent years. After being introduced to the United States in 2008, site traffic reached peaks of 15 million unique visits per month in 2011, roughly 20% of the traffic on Ebay at that time. Recent academic work has attempted to explain PAYGA outcomes, but standard static models with homogeneous agents consistently underpredict the number of total bids per auction observed in the data.[1][2]

[1]See, for example, Augenblick (2011) or Platt et al (2012).

[2]Overbidding has been documented for several other auction types as well. Examples include Kagel and Levin (1993), Murnighan (2002), Heyman et al (2004), and Lee and Malmendier (2011). However, as argued by Augenblick (2011), relative to PAYGA settings, these preceding studies either deal with items whose value is more difficult to observe or the magnitude of overbidding is much smaller.

This paper is the first to consider an explanation of the observed "overbidding" in PAYGA by relaxing the assumption that the game is played in a static environment. Using a new bid-level dataset with over 1,000 unique auctions and over 200,000 placed bids provided directly by a PAYGA hosting website, I examine a feature of these auctions that has received little attention in the existing literature, namely repeat encounters between the site's users across different auctions. Given that the identities of current and past bidders are publicly observable for active auctions, the scope for such repeated interactions exists. This environment allows users to learn about each other, and as a result, reputational strategies are plausible.

In particular, I find that there is a subset of users, which I call "power bidders", who participate in many more auctions and place many more bids in those auctions than the average user. I show that these aggressive users are able to form profitable reputations by strategically overbidding in auctions where many other participants and observers are present. After winning such an auction, power bidders' profits per auction increase by as much as \$18.51, which is very large compared to the loss per auction of \$5.20 earned by the average bidder. A similar effect is present for the user's winning percentage. I find that such a user's reputation affects others' participation decisions at the extensive margin. That is, fewer users choose to bid in auctions where the reputable power bidder is present, as opposed to players placing fewer bids on those auctions conditional on participation. This is consistent with the story that new, uninformed users arrive and bid against established bidders as usual, while experienced bidders choose to abstain. Interestingly, these effects are only observed if the power bidder wins after overbidding in an auction, as the profits and winning percentage of those who lose show no significant increase. This makes the aggressive bidding strategy both costly and risky, suggesting a potential explanation for why not all bidders adopt it. Finally, I show that in auctions where no power bidders are present, there is no evidence of overbidding relative to the static model.

Several other papers attempt to explain PAYGA's "excess" revenues by relaxing various assumptions of the baseline model but maintaining its static nature. Both Augenblick (2011) and Platt et al (2012) demonstrate that a standard static model with homogeneous, risk-neutral, and rational bidders underpredicts the number of bids placed in practice for PAYGA. The former asserts that bidders are irrational, and shows that a model with agents who obey a sunk cost fallacy and underestimate future regret can account for why they place more bids than a rational agent would. The latter concludes that the "excess" number of bids is due to agents' risk loving preferences, and as a result, classifies PAYGA as a form of gambling. Alternatively, Byers, Mitzenmacher, and Zervas (2010) consider disequilibrium behavioral explanations where heterogeneous agents are endowed with incorrect beliefs about their opponents and never learn. In contrast, I present evidence for an explanation involving rational agents, and I show that risk loving preferences are not an adequate explanation for my data.

Consistent with my findings, both Augenblick (2011) and Byers et al (2010) provide evidence that aggressive bidding strategies increase users' expected profits on average.[3] This supports the idea that aggressive bidding is rational in PAYGA. However, both of these papers only establish the positive impact on profits of aggressive bidding for *current* auctions, whereas I study how aggressive bidding in current auctions affects users' profits in *future* auctions. Augenblick (2011) hypothesizes that experienced players may form reputations across auctions, but it does not explore this idea empirically.[4] Furthermore, neither of these papers directly considers the possibility that the documented overbidding relative to the

[3]Augenblick (2011) shows that the expected profit of a bid increases with the number of bids a user has already placed in that auction. This approach has been criticized by Byers et al (2010) as being biased, since the hazard rate of an auction is generally increasing with the total number of bids placed. Byers et al (2010) shows that expected profit of a player is higher for auctions where that player bid more aggressively. In contrast to my work, both of these papers include how quickly a user bids as part of their definition of aggressive bidding.

[4]Augenblick (2011) notices that experience has a positive impact on the profitability of a bid, after controlling for the use of aggressive strategies. It speculates that reputation may be an omitted variable which affects profits and is correlated with experience, but no evidence is provided to support this hypothesis.

baseline model may be due to the presence of such reputation-based strategies.

The remainder of the paper is organized as follows. Section 2 provides a brief background on PAYGA and summarizes my data. Section 3 presents the baseline model and demonstrates its failure at matching the data. Finally, section 4 discusses the relevance of aggressive bidding strategies and reputation for explaining the data, and section 5 concludes.

4.2 Background and Data

4.2.1 The Basics of Pay-As-You-Go Auctions

PAYGA is a popular new online auction format. It was introduced in the United States by Swoopo.com in October, 2008, and since then the market has grown to several million unique visitors per month across dozens of websites.

In PAYGA, each auction starts with a price of $0 and a time limit. The auctions are observable by the general public, and at any time, an observer has the option of paying a fee to place a bid. Bidding takes place sequentially, and when a bid is placed, the price jumps up by an amount that is publicly known and fixed ahead of time for a given auction by the hosting website (the "bid increment"). That is, bidders are not able to choose the amount of their bid, as it is required to be equal to the previous bid plus the bid increment, which varies across auctions. Placing a bid also extends the time limit by several seconds, providing time for other users to observe the bid and potentially place another of their own. When the timer runs out, the current high bidder is given the option to purchase the item at their final bid price from the hosting website. By charging fees for each bid placed, the website is able to make money even though the auction winner almost always profits.

For example, a scenario in my sample is as follows. On Tuesday morning, the site posts an auction for a $50 gift card with a bid increment of $0.05 that is set to end at 7:00 pm that night. The starting price for the item is set at $0. At any time, users have the option of paying $0.40 to place a bid, which increases the gift card's price by $0.05 and lengthens the auction by 10 seconds. Bids are recorded sequentially and are publicly observable in real time. Any user may bid as many times as they like. After a total of 105 bids are placed, no other users elect to bid, so the timer runs out. As a result, the auction ends just after 7:17 pm, and the final bidder purchases the gift card for $5.25. The website collects $5.25 from the winning user and $42 in bid fees for a total loss of $2.75. Although the winning user made a profit, the website's losses were only small due to the large amount of collected bid fees.

PAYGA is similar to the more well known war of attrition (WOA).[5] The latter is a multiperiod game where at each stage, participants choose to either pay a fixed cost to bid or drop out of the game, and the last man standing wins. The former may also be thought of as a multistage game. The main distinctions are that in the former, there is only one bidder per period, and players may continue to bid in the future even if they did not bid in previous periods. As pointed out by Augenblick (2011), from a theoretical perspective, another important difference is that while in a WOA the value of the good is constant across periods, it declines over time in PAYGA, thereby eliminating the stationarity of WOA.

4.2.2 Data

My data are bid-level and provided directly by a PAYGA hosting website. They include observations from 1,092 auctions for 161 unique items with a total of 213,052 bids placed by 794 unique users. The observations in my sample include all bids in all auctions between September 2009, when the site commenced operations, and May 2010. The merchandise being sold ranged from cash equivalents, such as gift cards, to packs of bids to popular electronics, such as flat screen TVs and digital cameras. The bid increments ranged from $0.01 to $1.00. Smaller bid increments were much more common, however, as nearly 65% of

[5]Further applications and discussion of WOA can be found in Bliss and Nalebuff (1984), Fudenberg and Tirole (1986), and Alesina and Drazen (1991).

auctions had an increment of \$0.01 and almost 95% were less than \$0.20. The cost of bids was mostly constant at \$0.40, but it increased in early March 2010 to \$0.60. Although I do not observe individuals' values for the popular electronics, it is reasonable to assume the value of the gift cards and bid packs are equal to their face value. Also since only popular, new electronic products were offered, the market price for those items was well established, and I estimated their values as the market price listed on Amazon.com from the time they were being auctioned.

Overall the auctions in my sample are well described by the previous section, but a few additional details deserve mentioning. First, users had the option of using an automated bidding service where they could enter their maximum desired bid and maximum desired number of bids, and it would act accordingly. The automated bidding service was designed to mimic real users' behavior by placing bids at a random time within a certain proximity to the end of the auction, and who was using the service was not public information, though I do observe it in my data. Second, in practice, it is possible that multiple people could place bids at (almost) exactly the same time, so the perceived stage of the game could jump by more than one period. This issue is mitigated by the automated bidding service, which is able to perfectly observe the instant when another user bids. Third, due to internet connection latency, it is possible that someone intended to bid, but was unable to do so before the time ran out. As a result, some auctions were cut shorter than they otherwise would have been, but this only implies that any evidence of overbidding may be understated. Finally, bids had to be purchased in packs of at least 10, rather than individually, but since this is still a relatively small number, I ignore this complication in my analysis.

Obtaining data directly from an auctions website is advantageous for the study of reputation as it provides a complete history of every bid placed in every auction on the site during the given time period. This is in contrast to all previous work which uses data recorded ("scraped") from publicly available information. For example, Platt et al (2012) only observes auction level data, since they scraped information from an auctions summary page of a website, which omits information on the bidding behavior of everyone other than the winner.[6] Meanwhile, Augenblick (2011) was able to scrape bid-level data, but technical issues prevent him from capturing all auctions or all bids placed within any given auction. In contrast, my dataset has no measurement error, and therefore permits a more accurate study of individual users' behavior within and across auctions.

As a starting point, I present some basic auction-level summary statistics in Table 1. Here, I define normalized time as the total number of bids placed in an auction divided by the item's price. This provides a sense of the number of bids placed that is comparable across auctions for different items.[7] The majority of auctions had a penny bid increment, and auctions with lower increments tended to last longer. Focusing on normalized time, bid pack auctions were the longest and electronics the shortest. However, these numbers are associated with very large standard errors, and as a result, the differences in auction lengths across item types was not significant.

One of the most noteworthy features of the data is the large standard error associated with average normalized time. Given that time is bounded from below by zero, this suggests that most auctions were relatively short, but the distribution has a thick right tail. Figure 1 illustrates this feature by plotting the estimated kernel density for normalized time.[8] In fact, the median is only 1.32 while the 90% quantile is 6.9, and the maximum is 37.35. Therefore, any explanation of the data must address why a relatively small subset of auctions last so much longer than others, after controlling for the item's value.

[6]The data does include the last 10 bids placed in the auction, but this is a negligible portion of the total bids placed on average.

[7]The use of this normalized time measure was suggested by Augenblick (2011) who justifies it in the context of the baseline model, as I will discuss in section 3.2.

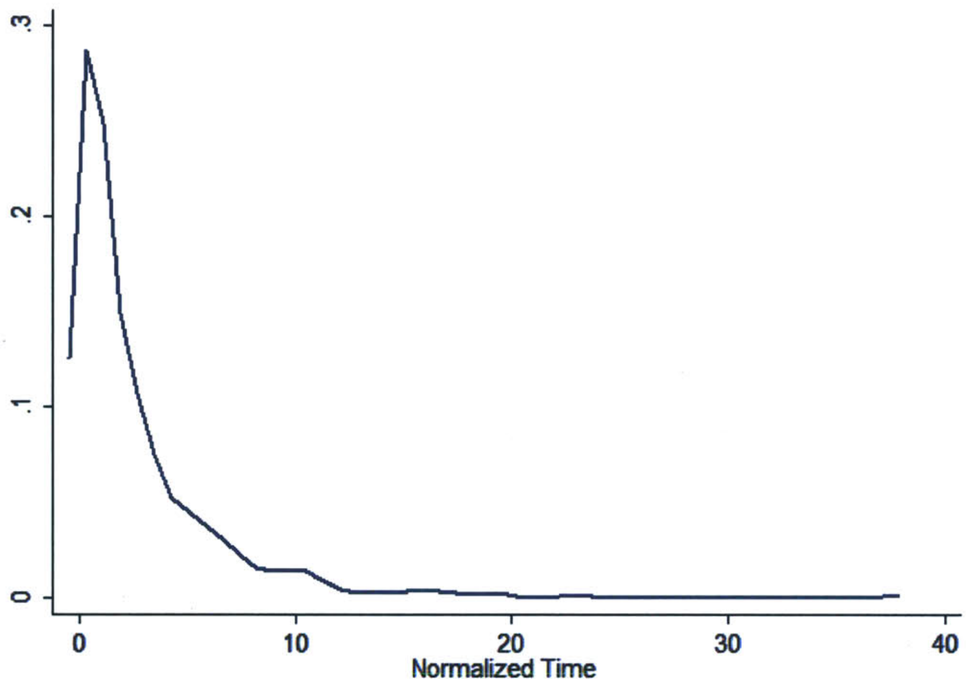
[8]Specifically, Figure 1 uses an Epanechnikov kernel with a bandwidth selected by Silverman's rule.

Table 1: Auction-Level Mean Statistics

	N	No. Bidders	Item Value	No. Bids	Norm. Time	Winner Profits
All	1,092	6.36	$78.58	195.10	2.65	$44.45
		(4.14)	(71.62)	(384.10)	(3.61)	(74.75)
Item Type:						
Bid Packs	390	5.67	$43.12	125.42	3.23	$22.57
		(2.99)	(22.47)	(160.96)	(3.91)	(28.76)
Electronics	150	8.79	$186.01	384.17	1.83	$116.47
		(6.42)	(102.32)	(645.91)	(2.63)	(122.18)
Gift Cards	506	6.33	$72.75	204.18	2.63	$37.44
		(3.84)	(48.65)	(395.63)	(3.67)	(67.85)
Bid Inc.:						
$0.01	709	6.62	$80.40	231.21	3.11	$41.86
		(3.76)	(67.47)	(413.36)	(4.05)	(79.66)
$0.05	88	8.42	$85.38	218.58	3.12	$53.05
		(6.88)	(101.30)	(406.78)	(3.32)	(80.73)
$0.10	53	5.75	$50.86	84.53	1.77	$31.80
		(3.39)	(38.82)	(106.33)	(1.88)	(31.93)
$0.15	94	4.79	$60.47	72.76	1.43	$39.90
		(3.10)	(50.40)	(105.75)	(1.48)	(45.62)

Notes: Standard errors are reported in parentheses.

Figure 1: Normalized Time Density



4.3 Baseline Model and Performance

4.3.1 Baseline Model

For the baseline model, I follow Augenblick (2011) and Platt et al (2012). Assume there are n identical, risk-neutral bidders vying to win an auction for a single good that is valued equally by all bidders at v. Time evolves discretely and is indexed by $t = \{0, 1, 2, 3, ...\}$.[9] In every period, each player must choose to either bid or not bid, and a period may only have at most one bidder. If some player bids in period t, they are denoted as the leader of that period, l_t, and the price is increased by the bid increment, $k > 0$. If no player bids in

[9]Although there is an auction timer in practice, this does not matter for the equilibrium of the baseline model, as bidders are assumed to have perfect attention.

period t, the auction ends, and the previous leader, l_{t-1}, becomes the winner. If more than one player chooses to bid, the auctioneer randomly selects which bid to accept. The player who bids in period t, l_t, pays a fixed bid cost, c. The final period of the auction where no bid is placed is denoted by $T+1$. The price begins at 0, so the price at the end of period $t < T+1$ is kt. Thus, the winning bid is kT, and the winner l_T earns $v - kT$ in profit, not counting their costs of bidding. Since the game is played over a relatively short period of time, it is assumed that bidders do not discount their consumption over the periods of the game. Finally, assume that $c < v - k$, so that bidding is possible in equilibrium.

In what follows, it will be useful to define the hazard function, $h(t)$, as the probability that t is the last period where a bid is placed given that the auction has reached period t. That is,

$$h(t) \equiv Pr\{t = T | t \leq T\}.$$

Assuming that $h(0), \ h(1) < 1$, Propositions 1-3 in Augenblick (2011) state that any equilibrium must have

$$h(t) = \begin{cases} \frac{c}{v-kt} & if\ 0 < t \leq \frac{v-c}{k} \\ 1 & else \end{cases}.$$

Thus, the model has specific predictions about auction hazard rates that can be compared with auction-level data.

The intuition for why the preceding hazard rates must occur in equilibrium is relatively simple. First, the assumption that $h(0), \ h(1) < 1$ rules out pathological cases, where there is no bidding or only one bid is ever placed.[10] Given that some bidding occurs, a natural guess for the hazard rates comes from simply equating a player's utility from bidding with their utility from not bidding, thereby making participants indifferent between bidding and

[10]See Augenblick (2011) for a further discussion of these pathological cases.

not bidding in any period. Doing so for a generic utility function produces,

$$h(t)\,u(w-c+v-kt)+(1-h(t))\,u(w-c)=u(w)$$

$$\Rightarrow$$

$$h(t) \;=\; \frac{u(w)-u(w-c)}{u(w-c+v-kt)-u(w-c)}$$

where w is the user's wealth level at time t. In the special case of risk-neutral agents, $u(w)=w$, and the expression simplifies to

$$h(t)=\frac{c}{v-kt}.$$

The final thing to notice is that nobody would bid if the cost of bidding is greater than the remaining net value of the item. This happens when $v-kt<c$, which can be rearranged to $t>\frac{v-c}{k}$. This is the final component of the equilibrium hazard rate derived formally by Augenblick (2011) and Platt et al (2012).[11]

Next I introduce survival rates, defined as the probability that an auction lasts more than t periods or

$$S(t)\equiv \Pr(t<T).$$

In order to solve for this quantity, it is convenient to work in continuous time. Since the intuition discussed above carries through to the continuous time case, I simply state the results from Augenblick (2011). In equilibrium, when time is continuous and under the same assumptions as before, hazard rates are as above, and the survival rates are given by

$$S(t)=e^{-\int_0^t h(s)ds},$$

[11] I refer the reader to either of these works for a complete listing of the necessary assumptions and formal proof that the described hazard rates must prevail in equilibrium.

which in the risk-neutral case becomes

$$S(t) = \left(1 - \frac{kt}{v}\right)^{\frac{c}{k}}$$

for $t < \frac{v}{k}$.

Finally, I derive the likelihood function. It can be computed from the hazard function as follows. Let $f(t)$ be the probability density function for auction ending times. Then

$$\begin{aligned} f(t) = Pr(t = T) &= Pr(t = T | t \leq T) Pr(t \leq T) \\ &= h(t) \prod_{s=1}^{t-1} [1 - h(s)]. \end{aligned}$$

The log-likelihood function therefore can be computed as,

$$\begin{aligned} \mathcal{L}(t_1, ..., t_N) &= log\left\{\prod_{n=1}^{N} h(t_n) \prod_{s=1}^{t_n-1} [1 - h(s)]\right\} \\ &= \sum_{n=1}^{N} log\left(h(t_n) \prod_{s=1}^{t_n-1} [1 - h(s)]\right) \\ &= \sum_{n=1}^{N} \left(log[h(t_n)] + \sum_{s=1}^{t_n-1} log[1 - h(s)]\right). \end{aligned}$$

With a given model for $h(.)$, one can estimate any free parameters in the standard way. This is not necessary for the baseline model because it does not employ any unobservable parameters. However, it will be useful when I consider extended versions of the model.

4.3.2 Model Performance on Auction Outcomes

In this section, I discuss how the baseline model's predictions fare against the auction-level data. As before, I define a normalized time measure as the number of bids placed divided by the value of the item, $\hat{t} = \frac{t}{v}$. This is following Augenblick (2011) who also derives the hazard and survival functions in terms of normalized time as

$$h\left(\hat{t}\right) = \frac{c}{1 - k\hat{t}}$$

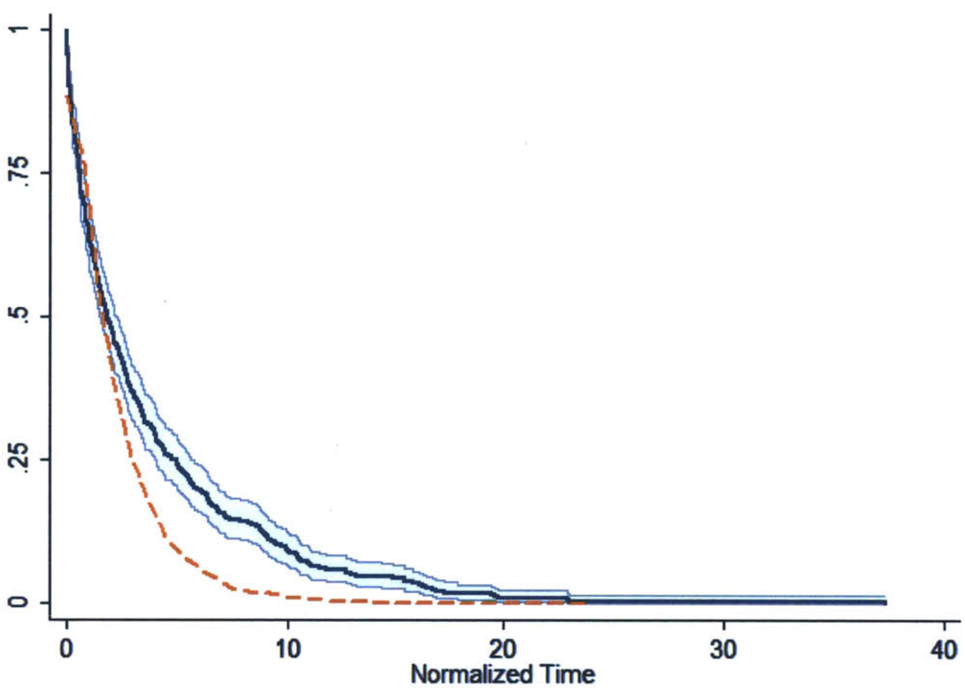
$$S\left(\hat{t}\right) = \left(1 - k\hat{t}\right)^{\frac{c}{k}}.$$

Using these expressions, I assess whether the model is capabale of accounting for the distribution of normalized auction lengths. The simplest way to compare the data with the model is to plot the hazard and survival functions. Since the hazard and survival functions from the model depend on both the cost of bids and the bid increment, a direct comparison to the data must hold those two parameters fixed. In order to mazimize sample size, Figure 2 plots the two functions for auctions with penny increments and a \$0.40 cost of bidding.[12] Panel A shows the Kaplan-Meier estimated survival function with 95% confidence bands[13] against that of the model. The model actually does reasonably well at capturing the hazard rates for shorter auctions, but a divergence occurs for those that lasted longer than roughly 2.5 normalized time units. In particular, longer auctions are significantly more likely in the data than the model would suggest. In other words, the model cannot explain why the distribution of auction lengths has such a thick right tail, as discussed in section 2. Panel B shows the smoothed hazard rates from the data against those from the model. The model predicts that the hazard rates should be slightly upward sloping, as $k > 0$ implies that the net value of the good is decreasing with each bid. Meanwhile, the hazard rates in the data appear relatively flat, or if anything, slighty downward sloping. Except for the ranges where the estimated standard errors are very large, the model's hazard rates are well above those observed in the data. Figure 2 confirms the findings of previous studies that there is overbidding in the data relative to the baseline model.

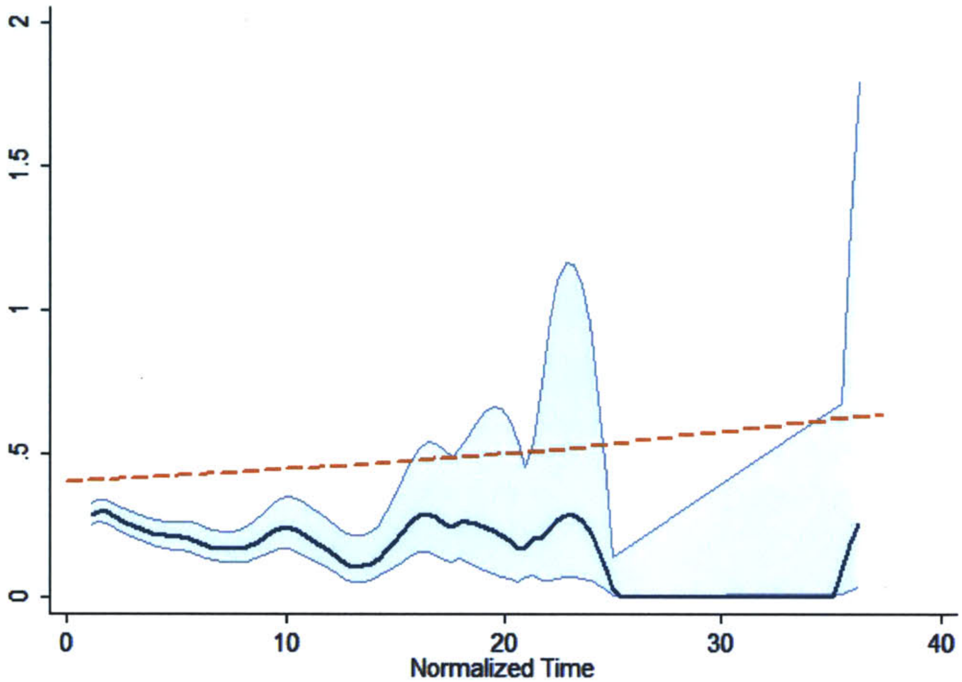
[12]The appendix contains the same plots for the period when the cost of bidding was \$0.60, which are qualitatively similar albeit less precisely estimated due to the smaller sample size.

[13]Confidence intervals are constructed from pointwise standard errors and assume normality of the estimates.

Figure 2: Survival and Hazard Rates vs. Baseline Model (Dashed)
Panel A: Survival Rates



Panel B: Hazard Rates



Notes: Observations include only those where the bid increment is \$0.01 and the bid price is \$0.40.

4.3.3 Entertainment Values / Bid Costs

Having reaffirmed the findings of previous work that the baseline model is a poor representation of the data, I turn now to examine whether there are relatively simple extensions that can rectify the discrepancy. First, I consider the possibility that bidders simply enjoy playing the game. The most straightforward way to model entertainment values is to assume that agents receive utility from each bid they place. This is in addition to the expected revenue from winning the auction. For risk-neutral agents, allowing for an entertainment value of bidding is equivalent to a uniform reduction in bid costs. Thus, given the likelihood function for the data, it is straightforward to estimate the entertainment value from bidding by maximum likelihood.

Letting e be the entertainment value that a participant receives from placing a bid, the hazard rate for this model becomes

$$h(t\ ;\ e) = \frac{c-e}{v-kt}.$$

Hence, I can perform maximum likelihood in a standard way to estimate the parameter e. The first line of Table 2 shows the estimated value along with standard errors. I find no significant evidence of any entertainment value in these auctions, although the point estimate was positive and fairly large relative to the cost of bidding.

4.3.4 Risk Loving Behavior

The next extension I consider is the one originally proposed by Platt et al (2010), namely that risk loving behavior by users leads them to overbid in auctions. The easiest way to model risk-loving behavior is with CARA utility because in that case the hazard rate does not depend on players' unobservable wealth level. To see this, one can plug a CARA utility function, $u(w) = \frac{1-e^{-\alpha w}}{\alpha}$, into the previous expression for the hazard rate. Doing so yields,

$$\begin{aligned} h(t) &= \frac{u(w) - u(w-c)}{u(w-c+v-kt) - u(w-c)} \\ &= \frac{1 - e^{-\alpha w} - \left(1 - e^{-\alpha(w-c)}\right)}{1 - e^{-\alpha(w-c+v-kt)} - \left(1 - e^{-\alpha(w-c)}\right)} \\ &= \frac{e^{\alpha c} - 1}{e^{\alpha c} - e^{-\alpha(-c+v-kt)}} \\ &= \frac{1 - e^{-\alpha c}}{1 - e^{-\alpha(v-kt)}}. \end{aligned}$$

Plugging this new hazard rate into the expression for the likelihood function from the previous section, I estimate the risk preference coefficient α. The results are shown in the second line of Table 2. The point estimate supports risk loving behavior, but again it is not signficant. The magnitude is also very small compared to those found in other contexts. For example, Jullien and Salanie (2000) estimate $\alpha = -0.06$ for horse race bettors, more than a full order of magnitude above what I find for PAYGA. Therefore, I conclude that risk-loving behavior alone is not sufficient to explain the data.

Table 2: Estimates of Model Parameters

Parameter	Estimate	Standard Error
e	0.13	0.09
α	-0.0022	0.0015

4.4 Reputation

Having been unsuccessful at explaining the discrepancy between the theory and the data within the confines of the static model, I turn to a new source of potential explanations which is the fact that bidders may encounter one another repeatedly across different auctions as well as within the same auction. In reality, users have access to a great deal of information about each other. Anytime a bid is placed, the bidder's identity is publicly displayed. Although this information is only available to those who are actively observing ongoing auctions, the site also makes public a full history of past auction winners and ending prices. Hence, it is reasonable that participants would use this information to better inform their own bidding strategies. This generates an entirely new set of strategies that may occur in equilibrium: those where bidders act in order to manipulate their own reputation to their advantage.

In this section, I consider the evidence for the plausibility and rationality of reputation-based strategies where users attempt to brand themselves as overly aggressive or unwilling to give up at any (or perhaps a very large) cost. If a bidder is able to garner such a reputation, it would have the potential to earn them profits in the future, as others become more reluctant to outbid them. Thus, such a strategy is potentially rational. On the other hand, establishing such a reputation may be quite costly, as one would need to convince others that they are willing to pay any price to win an auction. Whether or not it makes sense to do so in practice is an empirical question. Naturally, if a group of bidders is playing an aggressive strategy for reputational purposes, it could explain why the data exhibits a thick right tail in auction lengths, relative to the static model.

Conceptually, I assume that there are two types of bidders. One type, which I will refer to as "power bidders," plays very aggressively to establish a reputation. The other type consists of "regular" users who do not. There is constant user turnover as some people decide to no longer bid, perhaps after realizing it is not profitable for them, and new users setup accounts. When a player first arrives to the site, they do not know other users' types and must learn about them by participating in or observing active auctions. For simplicity, I assume that a power bidder's identity is only revealed after an extreme showing of aggression by that player, which I will refer to as a "reputation forming event." It is worth mentioning that although a single event may establish their reputation, it would not be a long-run equilibrium for a power bidder to stop playing aggressively afterwards because if that were optimal, others would be aware of it. Another natural question is whether it matters if a power bidder wins a reputation forming auction. Either way the user has shown the willingness to spend a large number of bids, but if they lose, it also signals that they are ultimately willing to give up. A formal model and equilibrium analysis of this setting is beyond the scope of the current paper and is left for future work.[14]

4.4.1 Descriptive Statistics

I begin by summarizing the user-level data in Table 3, which contains mean statistics for all users and power bidders separately. For power bidders, I consider two definitions of reputation forming events. The first is anytime a user places more than 500 bids in a single

[14] Hinosaar (2010) notes that the equilibrium in a model where users observe each other across auctions is generally asymmetric, and reputation-type equilibria can arise even without costs of reputation building.

auction. The second is any auction where a user spends more than 175% of the item's value on bids alone. I refer to the ratio of the amount spent on bids in an auction to the item's value as the "spending ratio." For the second definition, I only include items with a value of at least $45 because users are less likely to pay attention to someone who overspends on a small item. The cutoff for the second definition was chosen to produce roughly the same number of users as the first. I also later examine how the results change when other cutoffs are used. In general, I refer to a power bidder as a user who at any point in my sample has a reputation forming event. Thus, statistics for power bidders include observations (auctions) both before and after but not including the reputation forming events, which are excluded so as not to artificially inflate their average bids placed.

Table 3 reveals several interesting features of the data. Perhaps most importantly for my purposes, power bidders are very different from the average user. They participate in many more auctions and place many more bids in those auctions where they participate. Even though there are less than 20 power bidders, they account for more than 20% of all user-auction observations. They also place more than 3.5 times as many bids per auction as the average user. Hence, it is clear that the pool of auction participants is non-homogeneous, in contrast to the assumptions of the baseline model. Moreover, the definition of power bidder used herein does identify a set of users that is substantially more aggressive than the average user, even though it is based on activity observed in a single auction. Interestingly, the reported standard errors also suggest a great deal of heterogeneity within power bidders, which I will return to later.

Finally, I consider the scope for reputation formation through repeat encounters of other bidders implied by the summary statistics in Table 3. First, the large fraction of auction-user observations attributed to power bidders suggests that repeat auction observers or participants would have a good chance of encountering the same aggressive opponent multiple times. It is important to further point out that a user need not bid in an auction to observe it, and on average there were just as many users observing as participating. Second, the table shows the extremity of the reputation forming events I have chosen. Placing 500 bids in a single auction is nearly 20 standard deviations above the mean. This sort of behavior is likely to draw attention from observers. Not surprisingly, these auctions also tended to be during peak times for high profile items and with many participants. In the 19 auctions where at least one user placed more than 500 bids, there were 148 unique participants, and that does not include observers, since I am not able to identify them uniquely across auctions. Furthermore, those 148 bidders also tended to be repeat auction participants with an average of of 27.70 auctions played. Thus, if bidders are paying attention to each other, the reputation forming events I have chosen are likely to have had an impact.

Table 3: User-Level Mean Summary Statistics

User Sample:	All Users	Power Bidders	
Reputation Measure:	-	Placed 500 Bids	Spending Ratio 1.75*
No. Auctions	8.72	94.62	105.56
	(26.93)	(74.59)	(74.34)
Bids per Auction	13.97	59.72	52.40
	(24.43)	(20.95)	(22.85)
User Wins	1.37	34.94	31.83
	(8.67)	(39.36)	(38.25)
Winning %	2.59%	34.72%	28.51%
	(8.84%)	(25.07%)	(22.74%)
Profit per Auction	-$5.20	-$0.56	-$0.03
	(13.18)	(31.51)	(25.24)
N	792	16	18

*: To qualify, the spending ratio must be at least 1.75 on an auction where the item is worth at least $45.

4.4.2 Empirical Results

Given that a subset of users is much more aggressive than average and that reputation formation is plausible especially for these users, the next question I address is whether or not these power bidders are able to establish reputations that enhance their profitability. The basic empirical strategy that I employ is to look at power bidders' profits per auction before and after reputation forming events. I also consider the relevance of winning the auction where they placed the very large number of bids. The regression I run is

$$Profit_{i,j} = a + b_1 PBWin_{i,j} + b_2 PBLose_{i,j} + b_3 Exp_{i,j} + b_4 NBidders_j + b_5 v_j + b_6 FE_{i,j} + e_{i,j}.$$

The regression is at the auction-user level, and the dependent variable is the profits of user i in auction j. If auction j is a reputation forming event for any user, it is excluded. This is done because within those auctions, aggressive users may not be trying to maximize profits, since they serve the dual investment purpose of establishing their reputation and thereby generating more future profits.

The first two regressors are dummies that indicate whether or not user i has placed enough bids in a single auction to be defined as a power bidder prior to auction j. The first dummy, $PBWin_{i,j}$, indicates if the power bidder has won any of their previous reputation forming auctions, while the second indicates power bidders who have not yet won any of theirs. Thus, it is possible that a single user changes status from a power bidding loser to a power bidding winner, if they place the requisite number of bids multiple times. However, changing status from a winner to a loser is not possible. It is also impossible for either of these indicators to be equal to one for user i's first auction, since they only indicate power bidders who have already established their reputations in past auctions. This is different from the categories in Table 3, which included all auction observations from users who ever became established power bidders in my sample. The third regressor, $Exp_{i,j}$, is the number of auctions user i has placed a bid in prior to auction j. This captures a user's experience level, as it may be the case that players learn how to improve their strategies over time whether it be through general bidding tactics or learning about their opponents, and it controls for a more continuous form of learning than the discrete reputational events I have in mind.[15][16] The fifth and sixth regressors are the number of bidders and the price of the item in auction j. Both of these are likely to have a direct impact. They are also likely to be correlated with one another and with the presence of power bidders.

The final set of regressors are fixed effects to capture any unobservable factors. Although I have conceptually assumed there are two types of players, I do not want to rule out potential variation within each of these groups. Such variation may be due, for example, to user's intelligence, risk preferences, or PAYGA experience prior to joining the site I study. In order to control for such unobserved characteristics that do not vary over time, I include user fixed effects. I include month fixed effects to control for any changes in the site over time. For example, the site may increase advertising in one month, which could lead to heightened levels of competition in auctions and potentially lower user profits. Product category dummies are included for the four different categories of items sold: cash equivalents (e.g., gift cards), electronics, bid packs, and other. I also include dummies for the bid price, since it changed from \$0.40 to \$0.60 in the middle of a month, and the bid increment for auction j.

The results are shown in Table 4, and the estimates with the full set of controls are reported in columns 3 and 6 for the two definitions. Consistent with the hypothesis, the treatment effect of a power bidder establishing their reputation is positive and highly significant for the winners. The magnitude of the effect is also very large, either \$8.56 or \$18.51 depending on which definition is used, when one considers that the average player is losing \$5.20 per auction, and the average power bidder is roughly breaking even. Interestingly, the

[15]The results are qualitatively robust to using other measures of experience, including the number of bids placed in previous auctions and the number of past auction wins.

[16]Augenblick (2011) documents that over time bidders very slowly learn to become more profitable by bidding more aggressively.

effect is not the same for the losers, where the estimates are mixed between significantly negative and insignificant for the two definitions. Thus, I find that a power bidder must win a reputation forming auction before their reputation becomes a valuable asset.[17] This implies that aggressive bidders not only must pay a large cost to establish their reputation, but doing so is also risky. The coefficient on auction experience is small and insignificant, suggesting little evidence in favor of a more continuous learning process.

Table 4: Impact of Reputation on Users' Profits per Auction

Reputation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	-12.46***	-12.09**	-12.52**	4.62*	4.00	2.60
	(3.90)	(2.24)	(2.30)	(1.83)	(0.98)	(0.63)
Reputation + Win	22.71***	11.43**	8.56**	33.63***	20.07***	18.51***
	(7.74)	(2.29)	(2.18)	(7.59)	(2.90)	(2.61)
Auction Experience	-	-	0.02	-	-	-0.009
No. Bidders	-	-	-1.82***	-	-	-1.74***
Product Price	-	-	0.10***	-	-	0.07***
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
User FE	No	Yes	Yes	No	Yes	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	6,630	6,630	6,630	6,677	6,677	6,677
R-squared	0.0350	0.1453	0.1876	0.0268	0.1349	0.1684

Notes: The dependent variable is the profit of user i in auction j. The regression is run at the user-auction level, excluding reputation forming auctions. T-statistics are reported in parentheses. Standard errors are clustered by user.

[17] If winners and losers are pooled together, no significant effect is detected. These results are omitted to save space.

A similar pattern is observed when one uses the probability of winning an auction conditional on participation as the dependent variable, as shown in Table 5. The winning power bidders are very successful at establishing a useful reputation, as their probability of winning increases by 9% or 14% depending on the definition. Again, no clear pattern is present for the losers. The coefficient on experience is significant in this case, but the magnitude of the estimate is roughly 100 times smaller than the effect of forming a reputation through the discrete event.

Table 5: Impact of Reputation on Winning Probability

Reputation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	0.06***	-0.04	-0.02	0.20***	0.09**	0.11***
	(2.81)	(1.06)	(0.60)	(10.61)	(2.10)	(2.68)
Reputation + Win	0.38***	0.07**	0.09**	0.44***	0.10***	0.14***
	(17.73)	(2.15)	(2.34)	(14.42)	(3.47)	(3.90)
Auction Experience	-	-	-0.001***	-	-	-0.001***
No. Bidders	-	-	-0.01***	-	-	-0.008***
Product Price	-	-	6.7e-5	-	-	6.3e-5
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
User FE	No	Yes	Yes	No	Yes	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	6,630	6,630	6,630	6,677	6,677	6,677
R-squared	0.0832	0.2561	0.2711	0.0767	0.2598	0.2741

Notes: The dependent variable is an indicator for whether user i won auction j. The regression is run at the user-auction level, excluding reputation forming auctions. T-statistics are reported in parentheses. Standard errors are clustered by user.

Further diagnosing the precise impact of winning power bidders' reputations, I explore whether its source is the extensive or intensive margin or both. By extensive margin, I mean the number of bidders who choose to participate in auctions with the reputable winning power bidders. On the other hand, I define the intensive margin in this context as the number of bids placed against the winning power bidders by those who choose to participate against them. I proceed by measuring these two effects separately.

First, in order to measure the magnitude of the reputation effect at the intensive margin, I run another regression at the auction-user level. It is the same regression as the previous two except that the dependent variable is the number of bids placed by user i in auction j, and now the two reputation dummies indicate whether an established power bidder, winners and losers separately, are present in auction j. I will look separately at how reputation affects regular users and other power bidders.

Table 6 displays the results where only regular users are included. That is, I exclude all users who ever either won or lost a reputation forming event. The effect for winners on regular users at the intensive margin is negative for both definitions, as expected. However, in this case, it is only significant for the second definition where its magnitude is also large at 6.33, given that the average user only places 13.97 bids per auction. Not surprisingly, no significant effect is present for the losing power bidders. Table 7 shows the intensive margin for other power bidders, so it excludes all users who never won or lost a reputation forming event. The dummies in this case indicate when an established power bidder other than user i is present in auction j. The estimated effect is negative for the winners and moreso than the losers, but the estimates are not robustly significant.

Table 6: Number of Regular User Bids vs. Power Bidders

Repuation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	6.78***	6.63***	1.62	3.48***	1.14	-1.98
	(4.89)	(3.93)	(0.92)	(2.60)	(0.52)	(0.89)
Reputation + Win	1.54	-0.54	-1.76	-0.50	-3.78*	-6.33***
	(1.35)	(0.26)	(0.89)	(0.29)	(1.65	(2.74)
Auction Experience	-	-	-0.02	-	-	-0.10***
No. Bidders	-	-	1.40***	-	-	1.09***
Product Price	-	-	-0.001	-	-	0.03**
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
User FE	No	Yes	Yes	No	Yes	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	5,116	5,116	5,116	4,777	4,777	4,777
R-squared	0.0092	0.2230	0.2516	0.0018	0.2590	0.2933

Notes: The dependent variable is the number of bids placed by user i in auction j. The regression is run at the user-auction level, excluding reputation forming auctions and power bidders. T-statistics are reported in parentheses. Standard errors are clustered by user.

Table 7: No. Power Bidder Bids vs. Established Power Bidders						
Repuation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	23.13***	20.01***	7.33	9.26**	10.83**	-1.04
	(4.92)	(4.21)	(1.47)	(2.06)	(2.04)	(0.20)
Reputation + Win	4.36	1.21	-4.04	1.04	0.77	-10.33
	(1.07)	(0.25)	(0.84)	(0.17)	(0.10)	(1.26)
Auction Experience	-	-	-0.03	-	-	0.04
No. Bidders	-	-	3.78***	-	-	3.84***
Product Price	-	-	0.10*	-	-	0.298***
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
User FE	No	Yes	Yes	No	Yes	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	1,514	1,514	1,514	1,900	1,900	1,900
R-squared	0.0216	0.1001	0.1745	0.0022	0.1218	0.2085

Notes: The dependent variable is the number of bids placed by user i in auction j. The regression is run at the user-auction level, excluding reputation forming auctions and non-power bidders. T-statistics are reported in parentheses. Standard errors are clustered by user.

Next, I estimate the effect of reputation on the extensive margin. In order to do so, I use an auction-level regression where the dependent variable is the number of regular bidders who participate in an auction. The same controls are used as from the prior regressions, excepting those at the user-level. Table 8 presents the results. The effect of reputation at the extensive margin is significant, with fewer regular bidders choosing to bid against the winning power bidders. The estimated reduction in slightly more than one regular bidder per auction is large when compared with the average of just over six bidders per auction for the whole sample. Table 9 shows that the same pattern holds true for the number of power bidders in an auction. Since the average number of power bidders in an auction is 1.18, the estimated effects in Table 9 are economically large in magnitude as well.[18]

Altogether, I find that the positive effect of reputation on the profits for the winning power bidders comes primarily through the extensive margin, as fewer other users are willing to play against them. This makes sense given the constant inflow of new, uninformed users to the site. The pattern is consistent with the following story. Old users know not to bid against the established (winning) power bidders. This reduces the number of participants in auctions with established power bidders. However, new users are uninformed, so they enter and play the auctions as they would otherwise. As a result, conditional on participation, established power bidders' opponents can be expected to place the same number of bids.

[18]The results in Tables 8 and 9 are robust to including the experience level of the power bidders who are present. For example, in Table 8, adding two regressors to column 3, the maximum experiences of participating winning and losing power bidders, the effect of the presence of winning power bidders remains strongly significant at -0.86. This rules out the possibility that the results are due to power bidders gradually learning to pick auctions with fewer users. These results are omitted to save space.

Table 8: No. Regular Bidders vs. Power Bidders

Repuation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	1.04***	1.29***	1.00***	-0.54***	0.05	-0.07
	(3.46)	(3.95)	(3.27)	(2.23)	(0.81)	(0.23)
Reputation + Win	-1.55***	-0.81***	-1.06***	-1.28***	-0.48	-1.03***
	(7.11)	(2.96)	(4.06)	(4.66)	(1.51)	(3.59)
Product Price	-	-	0.02***	-	-	0.03***
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	1,073	1,073	1,073	1,067	1,067	1,067
R-squared	0.0420	0.1996	0.2933	0.0255	0.1937	0.3423

Notes: The dependent variable is the number of non-power bidders in auction j. The regression is run at the auction level, excluding reputation forming auctions. T-statistics are reported in parentheses. Standard errors are robust to heteroskedasticity.

Table 9: No. Power Bidders vs. Established Power Bidders						
Repuation Measure:	Placed 500 Bids			Spending Ratio 1.75		
Reputation + Lost	0.54***	0.54***	0.50***	1.22***	1.07***	0.98***
	(8.49)	(8.23)	(7.56)	(17.70)	(12.04)	(10.88)
Reputation + Win	-0.22***	-0.12**	-0.13**	-0.27***	-0.24**	-0.34***
	(5.56)	(2.25)	(2.35)	(3.64)	(2.49)	(3.62)
Product Price	-	-	0.001***	-	-	0.004***
Bid Price FE	No	No	Yes	No	No	Yes
Bid Inc FE	No	No	Yes	No	No	Yes
Month FE	No	Yes	Yes	No	Yes	Yes
Product Category FE	No	Yes	Yes	No	Yes	Yes
Observations	1,073	1,073	1,073	1,067	1,067	1,067
R-squared	0.1079	0.2284	0.2701	0.2307	0.3675	0.4247

Notes: The dependent variable is the number of power bidders in auction j. The regression is run at the auction level, excluding reputation forming auctions. T-statistics are reported in parentheses. Standard errors are robust to heteroskedasticity.

I turn now to measuring the effect that these aggressive users have on the aggregate overbidding in PAYGA, relative to the baseline model. First of all, through the reputation forming events themselves, there is already a large amount of overbidding taking place, by definition, and I have shown evidence that this is rational. However, it is worth going further to measure the impact of the power bidders on overbidding outside of the reputation forming auctions. To do so, I split the user-auction observations for power bidders into three groups. The first includes those who won a reputational auction in the past. The second is those who will win a reputational event in the future but have not yet. The third is those who place enough bids in a single auction to count as a power bidder but never win a reputational auction. I then regress the total number of bids in an auction on the number of each type of power bidders that are present, and the other controls that I have previously used. The results are in Table 10 where the observations exclude the reputational events. On the whole, the presence of power bidders has a significant increase on the number of bids placed in an auction, after controlling for the total number of bidders. Also consistent with the theory, the estimated effects are much larger for those users who have not yet won a reputational event as compared with those who have won a such an event in the past. In particular, the estimated effect for group 1 is significantly different from the estimated effect for group 3. Moreover, the difference between the effect for group 2 is not significantly different from the effect for group 3. Thus, as one would expect, the majority of the effect on overbidding comes from the power bidders who have not yet been able to establish themselves.

Table 10: Impact of Reputation on Total Bids

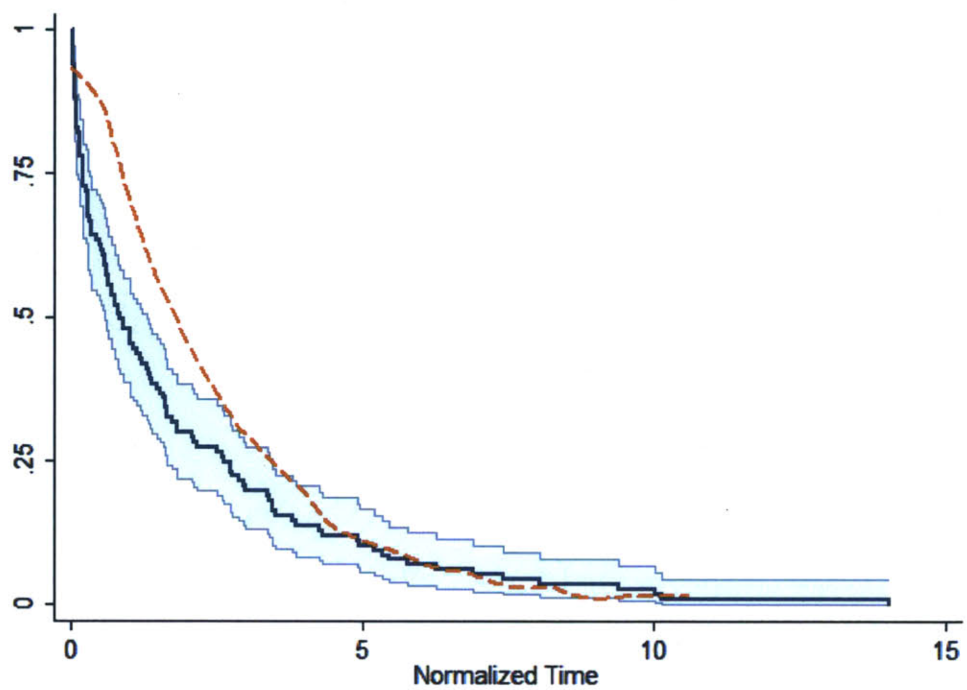
Repuation Measure:	Placed 500 Bids		Spending Ratio 1.75	
No. Rep. + Win (After) (1)	73.03***	25.99**	47.90***	6.10
	(8.15)	(2.48)	(2.89)	(0.26)
No. Rep. + Win (Before) (2)	79.64***	36.62***	107.36***	65.93***
	(6.92)	(3.41)	(4.55)	(2.85)
No. Rep. + Lost (Ever) (3)	131.55***	66.48***	104.88***	46.05***
	(9.74)	(4.74)	(8.30)	(3.99)
No. Bidders	-	32.93***	-	34.96***
Product Price	-	0.24	-	1.05***
Bid Price FE	No	Yes	No	Yes
Bid Inc FE	No	Yes	No	Yes
Month FE	No	Yes	No	Yes
Product Category FE	No	Yes	No	Yes
Observations	1,073	1,073	1,067	1,067
R-squared	0.2965	0.5458	0.2745	0.5671

Notes: The dependent variable is the number of bids in auction j. The regression is run at the auction level, excluding reputation forming auctions. T-statistics are reported in parentheses. Standard errors are robust to heteroskedasticity.

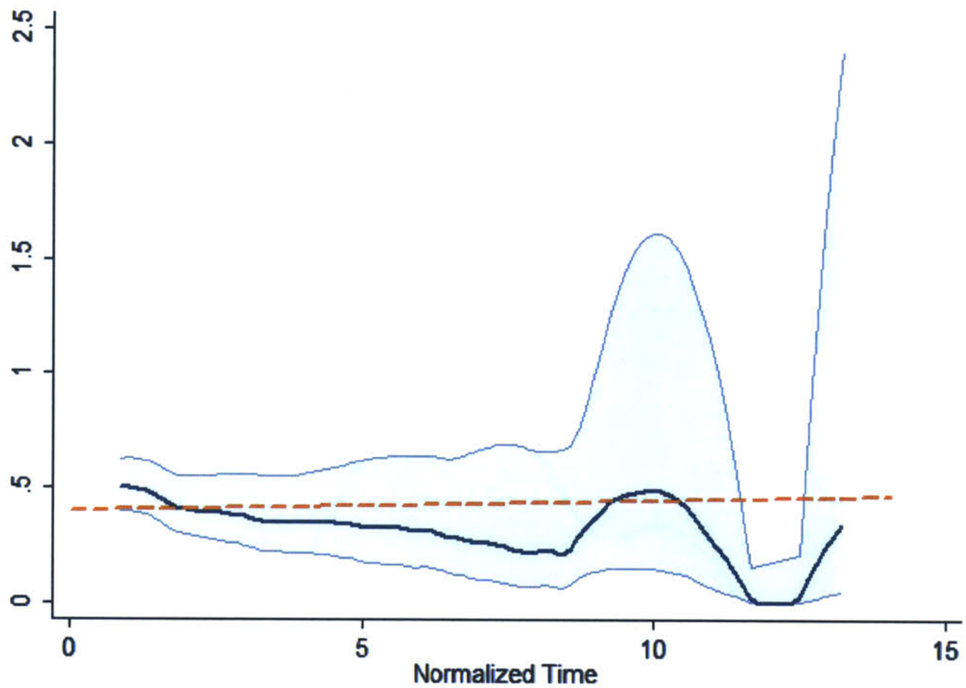
The last piece of evidence I show is a comparison between the baseline model and the data for auctions where no power bidders are present. Figure 3 reproduces the same hazard and survival rate plots as Figure 2 for this subsample of 117 observations. As before, I only include auctions with a bid price of $0.40 and a bid increment of $0.01. The pattern that emerges is very different from the previous figure. Whereas the full sample showed clear overbidding, the auctions without power bidders are quite well captured by the model. In fact, the survival rates for shorter auctions even exhibit some underbidding.

Figure 3: Survival and Hazard Rates vs. Baseline Model (Dashed)

Panel A: Survival Rates (Excluding Power Bidders)



Panel B: Hazard Rates (Excluding Power Bidders)



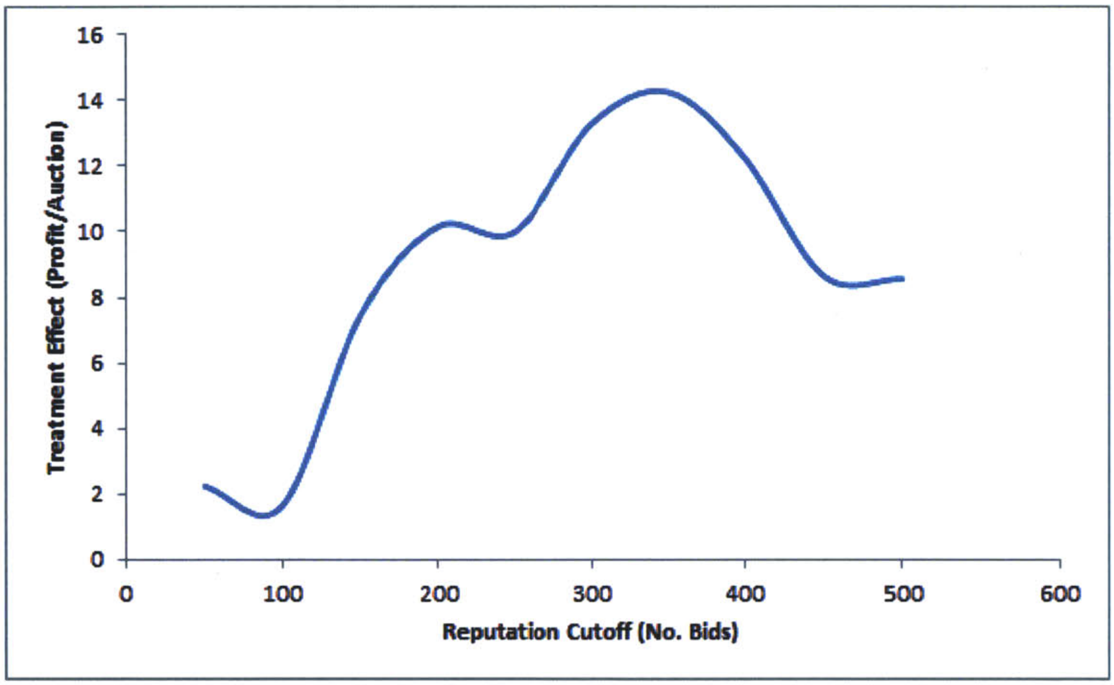
Notes: Observations include only those where the bid increment is \$0.01, the bid price is \$0.40 and no power bidder participated.

4.4.3 Robustness

Before concluding, I examine whether or not the results are robust to choosing different definitions for reputation forming events. Previously, I used 500 bid or a spending ratio of 1.75 in a single auction as the cutoff for defining a power bidder, and the events that established their reputations. Here, I considering varying this cutoff. Figure 4 shows the treatment effect of winning the reputation forming event on users' profits per auction across different cutoffs. In other words, I run the same regression in columns 3 and 6 of Table 4 for different cutoffs. Both panels show a clear upward slope as the cutoff is raised overall. In Panel A, there is a large gap between 100 and 200 bids placed in an auction, and in Panel B, a similar jump happens between a spending ratio of 1 and 1.1. This supports the assumption that learning occurs discontinuously in PAYGA. Although I omit the additional tables to save space, the main results of the previous section all hold when using a bid cutoff that is 200 or larger.[19]

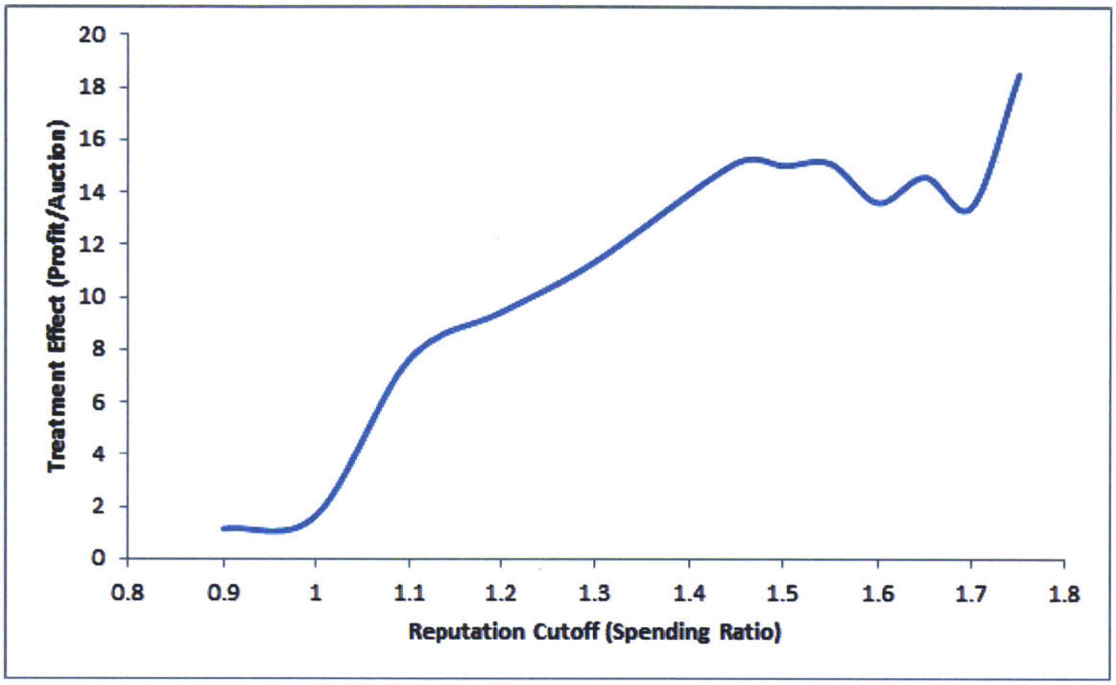
Figure 4: Treatment Effect by Cutoff

Panel A: Bid Cutoffs



[19]Specifically, I have tried using 200, 300, 400, and 500.

Panel B: Spending Ratio Cutoffs



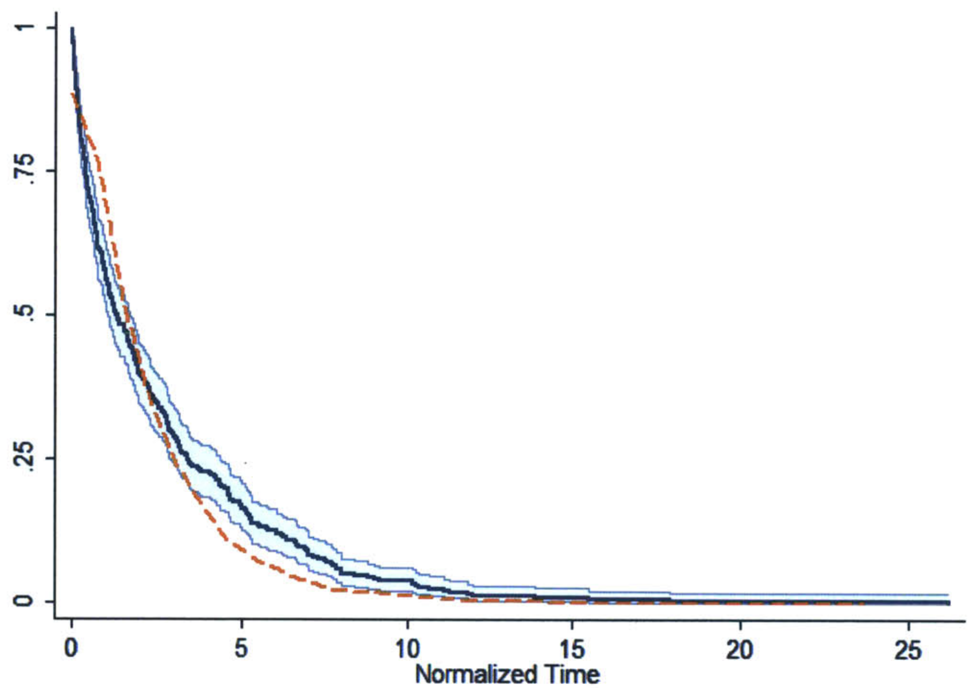
4.5 Conclusion

In this paper, I have shown that the apparent overbidding on PAYGA websites relative to a static baseline model is due to a small subset of users ("power bidders") who are much more aggressive than average. In contrast to the assumptions of the standard model, I showed that in practice there is scope for repeated interactions between the same users across different auctions. As a result, I argued that aggressive bidding may be rational, if those users are able to form a reputation that deters future competition. I presented evidence that power bidders are able to build reputations after an extreme amount of bidding in a single auction, as long as they also win the auction. The requirement that the power bidder must win the reputational auction makes the strategy risky in addition to costly, but I have shown that it does lead to significant increases in profitability and winning percentage, both with economically large magnitudes. Most of the gains to being aggressive come from the extensive margin, whereby fewer users choose to bid against a power bidder who has successfully established themselves. Considering only the auctions that do not contain any power bidders, no evidence of overbidding in PAYGA remains.

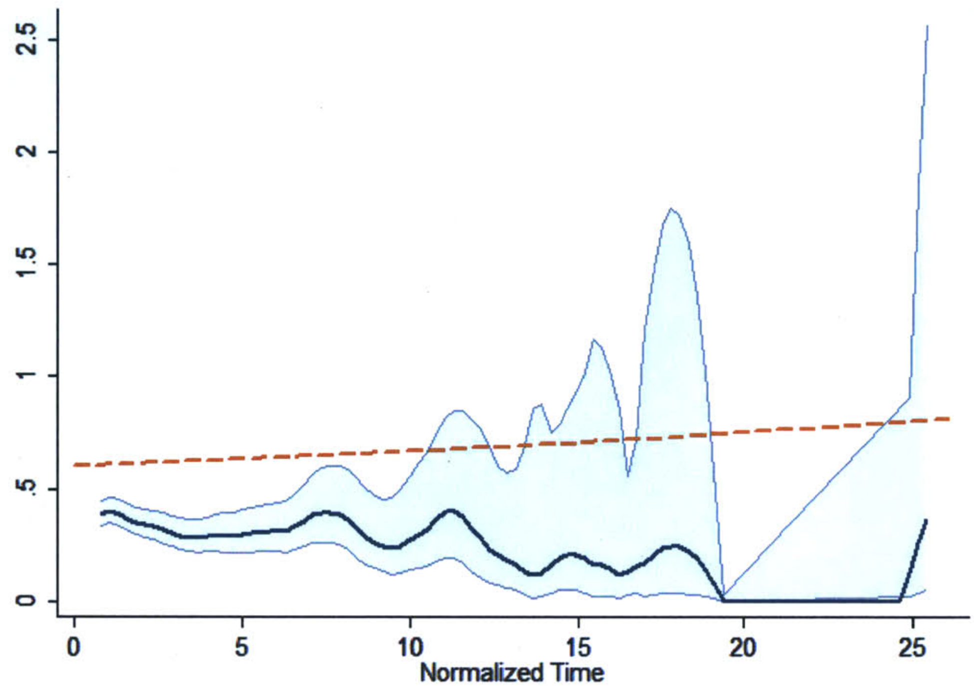
4.6 Appendix

Figure A1: Survival and Hazard Rates vs. Baseline Model (Dashed)

Panel A: Survival Rates



Panel B: Hazard Rates



Notes: Observations include only those where the bid increment is $0.01 and the bid price is $0.60.

Bibliography

[1] ABEL, A. An exploration of the effects of pessimism and doubt on asset returns. *Journal of Economic Dynamics and Control 26*, 7-8 (2002), 1075–1092.

[2] ADAM, K., AND MARCET, A. Internal rationality, imperfect market knowledge and asset prices. *Journal of Economic Theory 146*, 3 (2011), 1224–1252.

[3] ADAM, K., MARCET, A., AND NICOLINI, J. P. Stock market volatility and learning. Working Paper Series 862, European Central Bank, 2008.

[4] ALDRICH, E., AND GALLANT, A. R. Habit, long-run risks, prospect? a statistical inquiry. *Journal of Financial Econometrics 9*, 4 (2011), 589–618.

[5] ALESINA, A., AND DRAZEN, A. Why are stabilizations delayed? *American Economic Review 81*, 5 (1991), 1170–88.

[6] ALTIG, D., CHRISTIANO, L., EICHENBAUM, M., AND LINDÉ, J. Technical appendix to "Firm specific capital, nominal rigidities, and the business cycle". Manuscript, Northwestern University, 2004.

[7] ALTIG, D., CHRISTIANO, L., EICHENBAUM, M., AND LINDE, J. Firm-specific capital, nominal rigidities and the business cycle. Working Paper 11034, National Bureau of Economic Research, 2005.

[8] ALTIG, D., CHRISTIANO, L., EICHENBAUM, M., AND LINDÉ, J. Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics 14*, 2 (2011), 225–247.

[9] ANDERSON, E., HANSEN, L., AND SARGENT, T. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association 1*, 1 (2003), 68–123.

[10] AUGENBLICK, N. Consumer and producer behavior in the market for penny auctions: A theoretical and empirical analysis. Working paper, Stanford University, 2011.

[11] BACKUS, D., FORESI, S., AND TELMER, C. Affine term structure models and the forward premium anomaly. *The Journal of Finance 56*, 1 (2001), 279–304.

[12] Bai, J., and Ng, S. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25*, 1 (2007), 52–60.

[13] Bai, J., and Ng, S. Extremum estimation when the predictors are estimated from large panels. *Annals of Economics and Finance 9*, 2 (2008), 201–222.

[14] Bai, J., and Ng, S. Instrumental variable estimation in a data rich environment. *Econometric Theory 26*, 06 (2010), 1577–1606.

[15] Bakshi, G., and Skoulakis, G. Do subjective expectations explain asset pricing puzzles? *Journal of Financial Economics 98*, 3 (2010), 462–477.

[16] Bansal, R., Dittmar, R., and Kiku, D. Cointegration and consumption risks in asset returns. *Review of Financial Studies 22*, 3 (2009), 1343–1375.

[17] Bansal, R., Gallant, A. R., and Tauchen, G. Rational pessimism, rational exuberance, and asset pricing models. *The Review of Economic Studies 74*, 4 (2007), 1005–1033.

[18] Bansal, R., and Kiku, D. Cointegration and long-run asset allocation. *Journal of Business & Economic Statistics 29*, 1 (2011), 161–173.

[19] Bansal, R., Kiku, D., and Yaron, A. An empirical evaluation of the long-run risks model for asset prices. Working Paper 15504, National Bureau of Economic Research, 2009.

[20] Bansal, R., Kiku, D., and Yaron, A. Long run risks, the macroeconomy, and asset prices. *American Economic Review 100*, 2 (2010), 542–46.

[21] Bansal, R., and Yaron, A. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance 59*, 4 (2004), 1481–1509.

[22] Barro, R. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics 121*, 3 (2006), 823–866.

[23] Beeler, J., and Campbell, J. The long-run risks model and aggregate asset prices: An empirical assessment. Working Paper 14788, National Bureau of Economic Research, 2009.

[24] Bhansali, V. Volatility and the carry trade. *The Journal of Fixed Income 17*, 3 (2007), 72–84.

[25] Bils, M., and Klenow, P. Some evidence on the importance of sticky prices. *Journal of Political Economy 112*, 5 (2004), 947–985.

[26] Bliss, C., and Nalebuff, B. Dragon-slaying and ballroom dancing: The private supply of a public good. *Journal of Public Economics 25*, 1–2 (1984), 1–12.

[27] Boivin, J., and Giannoni, M. Has monetary policy become more effective? *The Review of Economics and Statistics 88*, 3 (2006), 445–462.

[28] Branch, W., and Evans, G. Intrinsic heterogeneity in expectation formation. *Journal of Economic Theory 127*, 1 (2006), 264–295.

[29] Branch, W., and Evans, G. Model uncertainty and endogenous volatility. *Review of Economic Dynamics 10*, 2 (2007), 207–237.

[30] Bray, M., and Savin, N. Rational expectations equilibria, learning, and model specification. *Econometrica 54*, 5 (1986), 1129–60.

[31] Brennan, M., and Xia, Y. Stock price volatility and equity premium. *Journal of Monetary Economics 47*, 2 (2001), 249–283.

[32] Brunnermeier, M., Nagel, S., and Pedersen, L. Carry trades and currency crashes. In *NBER Macroeconomics Annual 2008*, vol. 23. University of Chicago Press, Chicago, Il, 2009, pp. 313–347.

[33] Burnside, C. The forward premium is still a puzzle. Working Paper 13129, National Bureau of Economic Research, 2007.

[34] Burnside, C. Carry trades and risk. Working Paper 17278, National Bureau of Economic Research, 2011.

[35] Burnside, C., Eichenbaum, M., Kleshchelski, I., and Rebelo, S. Do peso problems explain the returns to the carry trade? *Review of Financial Studies 24*, 3 (2011), 853–891.

[36] Burnside, C., Eichenbaum, M., and Rebelo, S. Carry trade and momentum in currency markets. *Annual Review of Financial Economics 3*, 1 (2011), 511–535.

[37] Byers, J., Mitzenmacher, M., and Zervas, G. Information asymmetries in pay-per-bid auctions. *Electronic Commerce: Proceedings of the 11th ACM Conference on Electronic Commerce*, ACM: New York, 1–12.

[38] Cagetti, M., Hansen, L., Sargent, T., and Williams, N. Robustness and pricing with uncertain growth. *Review of Financial Studies 15*, 2 (2002), 363–404.

[39] Campbell, J., and Cochrane, J. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*, 2 (1999), 205–251.

[40] Campbell, J., and Cochrane, J. Explaining the poor performance of consumption-based asset pricing models. *Journal of Finance 55*, 6 (2000), 2863–2878.

[41] Campello, M., Chen, L., and Zhang, L. Expected returns, yield spreads, and asset pricing tests. *Review of Financial Studies 21*, 3 (2008), 1297–1338.

[42] CANOVA, F. *Methods for Applied Macroeconomic Research.* Princeton University Press, Princeton, NJ, 2007.

[43] CECCHETTI, S., LAM, P., AND MARK, N. Asset pricing with distorted beliefs: Are equity returns too good to be true? *American Economic Review 90*, 4 (2000), 787–805.

[44] CHARI, V., KEHOE, P., AND MCGRATTAN, E. Are structural VARs with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics 55*, 8 (2008), 1337–1352.

[45] CHEN, X., AND LUDVIGSON, S. Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics 24*, 7 (2009), 1057–1093.

[46] CHRISTIANO, L., EICHENBAUM, M., AND EVANS, C. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*, 1 (2005), 1–45.

[47] CHRISTIANO, L., EICHENBAUM, M., AND VIGFUSSON, R. Assessing structural VARs. *NBER Macroeconomics Annual 21* (2007), 1–72.

[48] CLARIDA, R., DAVIS, J., AND PEDERSEN, N. Currency carry trade regimes: Beyond the fama regression. *Journal of International Money and Finance 28*, 8 (2009), 1375–1389.

[49] COCHRANE, J. *Asset Pricing.* Princeton University Press, Princeton, NJ, 2001.

[50] COGLEY, T., AND NASON, J. Output dynamics in real-business-cycle models. *American Economic Review 85*, 3 (1995), 492–511.

[51] COGLEY, T., AND SARGENT, T. Anticipated utility and rational expectations as approximations of bayesian decision making. *International Economic Review 49*, 1 (2008), 185–221.

[52] COGLEY, T., AND SARGENT, T. The market price of risk and the equity premium: A legacy of the great depression? *Journal of Monetary Economics 55*, 3 (2008), 454–476.

[53] COGLEY, T., AND SARGENT, T. Diverse beliefs, survival and the market price of risk. *Economic Journal 119*, 536 (2009), 354–376.

[54] COIBION, O., AND GORODNICHENKO, Y. Information rigidity and the expectations formation process: A simple framework and new facts. Working Paper 16537, National Bureau of Economic Research, 2010.

[55] COLLARD, F., MUKERJI, S., SHEPPARD, K., AND TALLON, J.-M. Ambiguity and the historical equity premium. Economics Series Working Paper 550, University of Oxford, Department of Economics, 2011.

[56] CONSTANTINIDES, G., AND GHOSH, A. Asset pricing tests with long-run risks in consumption growth. *Review of Asset Pricing Studies 1*, 1 (2011), 96–136.

[57] DICECIO, R. Comovement: it's not a puzzle. Working Paper 2005-035, Federal Reserve Bank of St. Louis, 2005.

[58] DICECIO, R., AND NELSON, E. An estimated DSGE model for the united kingdom. *Federal Reserve Bank of St. Louis Review 89*, 4 (2007), 215–231.

[59] DRIDI, R., GUAY, A., AND RENAULT, E. Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics 136*, 2 (2007), 397–430.

[60] DUPOR, B., HAN, J., AND TSAI, Y.-C. What do technology shocks tell us about the new keynesian paradigm? *Journal of Monetary Economics 56*, 4 (2009), 560–569.

[61] EICHENBAUM, M., AND FISHER, J. Estimating the frequency of price re-optimization in calvo-style models. *Journal of Monetary Economics 54*, 7 (2007), 2032–2047.

[62] ENGEL, C. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance 3*, 2 (1996), 123–192.

[63] FAMA, E. Forward and spot exchange rates. *Journal of Monetary Economics 14*, 3 (1984), 319–338.

[64] FAMA, E., AND FRENCH, K. The cross-section of expected stock returns. *Journal of Finance 47*, 2 (1992), 427–65.

[65] FAMA, E., AND FRENCH, K. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 1 (1993), 3–56.

[66] FARHI, E., FRAIBERGER, S., GABAIX, X., RANCIERE, R., AND VERDELHAN, A. Crash risk in currency markets. Working Paper 15062, National Bureau of Economic Research, 2009.

[67] FILLAT, J., AND GARDUNO, H. GMM estimation of an asset pricing model with habit persistence. Working paper, University of Chicago, 2005.

[68] FISHER, J. Technology shocks matter. *FRB of Chicago Working Paper No. 2002-14* (2002).

[69] FUDENBERG, D., AND TIROLE, J. A theory of exit in duopoly. *Econometrica 54* (1986), 943–960.

[70] FUSTER, A., LAIBSON, D., AND MENDEL, B. Natural expectations and macroeconomic fluctuations. *The Journal of Economic Perspectives 24*, 4 (2010), 67–84.

[71] GABAIX, X. Variable rare disasters: A tractable theory of ten puzzles in macro-finance. *American Economic Review: Papers & Proceedings 98*, 2 (2008), 64–67.

[72] Gabaix, X. Disasterization: A simple way to fix the asset pricing properties of macroeconomic models. *American Economic Review 101*, 3 (2011), 406–09.

[73] Gilboa, I. Expectation and variation in multi-period decisions. *Econometrica 57*, 5 (1989), 1153–69.

[74] Golosov, M., and Lucas Jr, R. Menu costs and phillips curves. *Journal of Political Economy 115*, 2 (2007), 171–199.

[75] Gourieroux, C., Monfort, A., and Renault, E. Indirect inference. *Journal of Applied Econometrics 8*, S1 (1993), S85–S118.

[76] Gourio, F., Siemer, M., and Verdelhan, A. International risk cycles. Working Paper 17277, National Bureau of Economic Research, 2011.

[77] Goyal, A., and Welch, I. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*, 4 (2008), 1455–1508.

[78] Guvenen, F. A parsimonious macroeconomic model for asset pricing. *Econometrica 77*, 6 (2009), 1711–1750.

[79] Hall, A., Inoue, A., Nason, J., and Rossi, B. Information criteria for impulse response function matching estimation of DSGE models. Working Paper 10-28, Duke University, Department of Economics, 2010.

[80] Hamilton, J. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

[81] Hansen, L. Large sample properties of generalized method of moments estimators. *Econometrica 50*, 4 (1982), 1029–1054.

[82] Hansen, L., Heaton, J., and Li, N. Consumption strikes back? measuring long-run risk. *Journal of Political Economy 116*, 2 (2008), 260–302.

[83] Hansen, L., Sargent, T., and Tallarini, T. Robust permanent income and pricing. *Review of Economic Studies 66*, 4 (1999), 873–907.

[84] Hansen, L., Sargent, T., and Wang, N. Robust permanent income and pricing with filtering. *Macroeconomic Dynamics 6*, 1 (2002), 40–84.

[85] Hansen, L., and Singleton, K. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica 50*, 5 (1982), 1269–86.

[86] Heyman, J., Orhun, Y., and Ariely, D. Auction fever: The effect of opponents and quasi-endowment on product valuations. *Journal of Interactive Marketing 18*, 4 (2004), 7–21.

[87] Hinnosaur, T. Penny auctions are unpredictable. Working paper, Northwestern University, 2010.

[88] Iacoviello, M. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review 95*, 3 (2005), 739–764.

[89] Jensen, M., Scholes, M., and Black, F. The capital asset pricing model: Some empirical tests. In *Studies in the Theory of Capital Markets*. Praeger Publishers Inc., New York, NY, 1972, p. 79–121.

[90] Jorda, O. Estimation and inference of impulse responses by local projections. *American Economic Review 95*, 1 (2005), 161–182.

[91] Jordà, O., and Kozicki, S. Estimation and inference by the method of projection minimum distance. Working Paper 07-56, Bank of Canada, 2007.

[92] Jullien, B., and Salanie, B. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy 108*, 3 (2000), 503–530.

[93] Jurek, J. Crash-neutral currency carry trades. Working paper, Princeton University, 2009.

[94] Kagel, J., and Levin, D. Independent private value auctions: Bidder behaviour in first-, second- and third-price auctions with varying numbers of bidders. *Economic Journal 103*, 419 (1993), 868–79.

[95] Kehoe, P. Assessing structural VARs: comment. *NBER Macroeconomics Annual 21* (2007), 73–96.

[96] Kilian, L., and Kim, Y. Do local projections solve the bias problem in impulse response inference? CEPR Discussion Paper 7266, Center for Economic and Policy Research, 2009.

[97] Klenow, P., and Kryvtsov, O. State-dependent or time-dependent pricing: Does it matter for recent US inflation? *The Quarterly Journal of Economics 123*, 3 (2008), 863–904.

[98] Kocherlakota, N. The equity premium: It's still a puzzle. *Journal of Economic Literature 34*, 1 (1996), 42–71.

[99] Kydland, F., and Prescott, E. Time to build and aggregate fluctuations. *Econometrica 50*, 6 (1982), 1345–1370.

[100] Kydland, F., and Prescott, E. Hours and employment variation in business cycle theory. *Economic Theory 1*, 1 (1991), 63–81.

[101] Lawrence, E., Geppert, J., and Prakash, A. An empirical investigation of the campbell-cochrane habit utility model. *Journal of Business Finance & Accounting 36*, 5-6 (2009), 774–791.

[102] Lettau, M., and Ludvigson, S. Euler equation errors. Working Paper 11606, National Bureau of Economic Research, 2005.

[103] Lettau, M., and Ludvigson, S. Euler equation errors. *Review of Economic Dynamics 12*, 2 (2009), 255–283.

[104] Ljungqvist, L., and Uhlig, H. Optimal endowment destruction under campbell-cochrane habit formation. Working Paper 14772, National Bureau of Economic Research, 2009.

[105] Longstaff, F., and Piazzesi, M. Corporate earnings and the equity premium. *Journal of Financial Economics 74*, 3 (2004), 401–421.

[106] Ludvigson, S., and Ng, S. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics 83*, 1 (2007), 171–222.

[107] Ludvigson, S., and Ng, S. Macro factors in bond risk premia. *Review of Financial Studies 22*, 12 (2009), 5027 –5067.

[108] Lustig, H., Roussanov, N., and Verdelhan, A. Countercyclical currency risk premia. Working Paper 16427, National Bureau of Economic Research, 2010.

[109] Lustig, H., Roussanov, N., and Verdelhan, A. Common risk factors in currency markets. *Review of Financial Studies 24*, 11 (2011), 3731–3777.

[110] Lustig, H., and Verdelhan, A. The cross section of foreign currency risk premia and consumption growth risk. *American Economic Review 97*, 1 (2007), 89–117.

[111] Lustig, H., and Verdelhan, A. The cross-section of foreign currency risk premia and consumption growth risk: A reply. Working Paper 13812, National Bureau of Economic Research, 2008.

[112] Lütkepohl, H. Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process. *Econometric Theory 4*, 1 (1988), 77–85.

[113] Lütkepohl, H., and Poskitt, D. Estimating orthogonal impulse responses via vector autoregressive models. *Econometric Theory 7*, 4 (1991), 487–496.

[114] Lutkepohl, H., and Saikkonen, P. Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics 81*, 1 (1997), 127–157.

[115] Malmendier, U., and Lee, Y. The bidder's curse. *The American Economic Review 101*, 2 (2011), 749–787.

[116] Mehra, R., and Prescott, E. The equity premium: A puzzle. *Journal of Monetary Economics 15*, 2 (1985), 145–161.

[117] Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A. Carry trades and global foreign exchange volatility. CEPR Discussion Paper 8291, Center for Economic and Policy Research, 2011.

[118] Menzly, L., Santos, T., and Veronesi, P. Understanding predictability. *Journal of Political Economy 112*, 1 (2004), 1–47.

[119] Murnighan, J. K. A very extreme case of the dollar auction. *Journal of Management Education 26*, 1 (2002), 56–69.

[120] Newey, W., and West, K. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 3 (1987), 703–708.

[121] Platt, B., Price, J., and Tappen, H. Pay-to-bid auctions. Working paper, Brigham Young University, 2012.

[122] Rabanal, P., and Rubio-Ramírez, J. Comparing new keynesian models of the business cycle: A bayesian approach. *Journal of Monetary Economics 52*, 6 (2005), 1151–1166.

[123] Rietz, T. The equity risk premium a solution. *Journal of Monetary Economics 22*, 1 (1988), 117–131.

[124] Rotemberg, J., and Woodford, M. An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual 12* (1997), 297–346.

[125] Santos, T., and Veronesi, P. Labor income and predictable stock returns. *Review of Financial Studies 19*, 1 (2006), 1–44.

[126] Savage, L. *The Foundations of Statistics.* John Wiley & Sons, Inc., New York, NY, 1954.

[127] Schorfheide, F. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics 15*, 6 (2000), 645–670.

[128] Shanken, J. On the estimation of beta-pricing models. *Review of Financial Studies 5*, 1 (1992), 1–55.

[129] Shapiro, M., and Watson, M. Sources of business cycle fluctuations. *NBER Macroeconomics Annual 3* (1988), 111–148.

[130] Sims, C. Models and their uses. *American Journal of Agricultural Economics 71*, 2 (1989), 489–494.

[131] Smets, F., and Wouters, R. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association 1*, 5 (2003), 1123–1175.

[132] STOCK, J., AND WATSON, M. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 460 (2002), 1167–1179.

[133] STOCK, J., AND WATSON, M. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*, 2 (2002), 147–62.

[134] STOCK, J., AND WATSON, M. Forecasting with many predictors. In *Handbook of Economic Forecasting*, vol. 1. Elsevier B.V., Amsterdam, The Netherlands, 2006, pp. 515–554.

[135] TALLARINI, T., AND ZHANG, H. External habit and the cyclicality of expected stock returns. *The Journal of Business 78*, 3 (2005), 1023–1048.

[136] TIMMERMANN, A. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics 108*, 4 (1993), 1135–45.

[137] TIMMERMANN, A. Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *The Review of Economic Studies 63*, 4 (1996), 523–557.

[138] URIBE, M., AND YUE, V. Country spreads and emerging countries: Who drives whom? *Journal of International Economics 69*, 1 (2006), 6–36.

[139] VERONESI, P. Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies 12*, 5 (1999), 975–1007.

[140] VERONESI, P. The peso problem hypothesis and stock market returns. *Journal of Economic Dynamics and Control 28*, 4 (2004), 707–725.

[141] WEIL, P. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics 24*, 3 (1989), 401–421.

[142] WEITZMAN, M. Subjective expectations and asset-return puzzles. *American Economic Review 97*, 4 (2007), 1102–1130.

[143] ZHANG, Y. Fluctuations of real interest rates and business cycles. *Annals of Economics and Finance 11*, 1 (2010), 185–208.