



Essays in Macroeconomics

Citation

Schaab, Andreas. 2021. Essays in Macroeconomics. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37368359>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the Committee for the
PhD in Business Economics have examined a dissertation
entitled

Essays in Macroeconomics

Presented by **Andreas Schaab**

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

A handwritten signature in black ink that reads 'Xavier Gabaix'.

Signature

Xavier Gabaix, Co-Chair

A handwritten signature in blue ink that reads 'Matteo Maggiori'.

Signature

Matteo Maggiori, Co-Chair

A handwritten signature in black ink that reads 'Ludwig Straub'.

Signature

Ludwig Straub

Date: __April 15, 2021__

Essays in Macroeconomics

A dissertation presented

by

Andreas Schaab

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2021

© 2021 Andreas Schaab

All rights reserved.

Dissertation Advisors:

Professor Xavier Gabaix
Professor Matteo Maggiori

Author:

Andreas Schaab

Essays in Macroeconomics

Abstract

My dissertation presents essays on topics in macroeconomics. In Chapter 1, I study the interplay between micro and macro uncertainty. I show that its effect on unemployment risk is the dominant transmission channel of macro uncertainty. Accounting for this channel, the overall effect of a macro uncertainty shock on activity is 2-3 times larger. In Chapter 2 which is joint work with Allen Zhang, we develop an adaptive sparse grid method for solving continuous-time dynamic programming problems. While standard finite-difference discretization leads to inconsistent solutions of differential equations on irregular grids, we show that a generalized sparse finite-difference method restores consistency. Our paper is accompanied by a general-purpose toolbox for solving partial differential equations on adaptive sparse grids. In Chapter 3, which is joint work with Christopher Clayton, we develop a theory of dynamic inflation targets. We study a dynamic mechanism design problem between a government (principal) and a central bank (agent). A “dynamic inflation target” implements the constrained efficient inflation level: the central bank reports its target one period in advance, with a linear incentive scheme for deviations from the target.

Contents

Abstract	iii
Acknowledgments	ix
Dedication	x
1 Micro and Macro Uncertainty	1
1.1 Introduction	1
1.2 Illustrative Two-Period Example	11
1.3 A HANK Model with Micro and Macro Uncertainty	16
1.3.1 Households	16
1.3.2 Firms	20
1.3.3 Capital producer	21
1.3.4 Nominal wage stickiness	22
1.3.5 Government policy and ZLB	23
1.3.6 Aggregation and market clearing	23
1.4 A Global Solution Method	25
1.4.1 Finite-dimensional distribution representations	28
1.4.2 How to overcome the curse of dimensionality?	31
1.4.3 How to find a consistent law of motion for α_t ?	33
1.4.4 How to Choose Efficient Representations $F(\cdot)$?	36
1.5 Taking the Model to the Data	37
1.5.1 Calibration strategy	38
1.5.2 The ins and outs of unemployment	40
1.5.3 Business cycle fluctuations	42
1.6 Micro and Macro Uncertainty	44
1.6.1 Transmission mechanism of macro uncertainty	45
1.6.2 Endogenous uncertainty spikes during crises	53
1.6.3 Uncertainty Multiplier	56
1.6.4 Macro uncertainty in the data	57
1.7 ZLB Spells and the Paradox of Thrift	60
1.7.1 ZLB spells with micro and macro uncertainty	60
1.7.2 Paradox of thrift revisited	60

1.8	Cost of Business Cycles	63
1.9	Conclusion	64
2	Using Adaptive Sparse Grids to Solve Heterogeneous-Agent Models in Continuous Time	65
2.1	Introduction	65
2.2	High-Dimensional Partial Differential Equations	69
2.2.1	Sparse Grid Construction	72
2.2.2	Function Representations in the Hierarchical Basis	76
2.2.3	Hierarchization Operators	83
2.3	Value Function Iteration on Adaptive Sparse Grids	85
2.3.1	Consistent Finite-Difference Schemes on Adaptive Sparse Grids	87
2.3.2	Efficient Operator Construction	88
2.3.3	Boundary Conditions	89
2.3.4	Grid Adaptation	92
2.3.5	Algorithm Structure	93
2.4	Krusell-Smith	94
2.4.1	The Krusell and Smith (1998) Model	94
2.4.2	High-Dimensional Distribution Approximation	97
2.4.3	Numerical Results	98
2.5	Quality Ladders Under Dynamic Oligopoly	101
2.5.1	Model	102
2.5.2	Numerical Results	103
2.6	Conclusion	104
3	A Theory of Dynamic Inflation Targets	105
3.1	Introduction	105
3.2	Model	107
3.2.1	Constrained Efficient Allocation	110
3.2.2	Mechanism Structure	112
3.2.3	Incentive Compatibility, Time Consistency, and Information	112
3.2.4	Naive Target Adjustment	114
3.3	Dynamic Inflation Target	115
3.3.1	Evolution of the Target	119
3.3.2	On the Nature of Inflation Targets	121
3.3.3	Target Adjustment in Practice	121
3.3.4	Relation to Conservative Central Banker	122
3.4	Inflation Targeting and Reputation	123
3.4.1	Social Costs of Reputation	124

3.5	Second-Best Mechanisms with Costly Transfers	127
3.5.1	Model	127
3.5.2	The Second-Best Allocation Rule	129
3.5.3	Sufficient Statistics for the Optimal Mechanism	132
3.5.4	Reversion to Dynamic Inflation Target	133
3.5.5	Second best with Average Transfers	134
3.5.6	Discussion	135
3.6	Conclusion	135
References		136
Appendix A Appendix to Chapter 1		145
A.1	Appendix for Section 1.3: Quantitative Model	145
A.1.1	Household Problem	146
A.1.2	Nexus between the HJB and KF Equations	155
A.1.3	Inflation Risk	157
A.1.4	Death Process and Blanchard (1985) Annuity Markets	159
A.1.5	Illustration of Walras' Law	162
A.2	Appendix for Section 1.4: Solution Method	164
A.2.1	Derivation of Stochastic KF Equation (Diffusion Processes)	164
A.2.2	Proof of Proposition 3 (Diffusion Processes)	168
A.2.3	Choosing $F(\cdot)$ from a parametric family	172
A.2.4	Non-parametric algorithm	173
A.3	Appendix for Section 1.5: Data and Empirics	176
A.3.1	Employment Transitions in CPS Micro Data	176
A.3.2	Business Cycle Moments	179
A.3.3	Macroeconomic Uncertainty Indices	179
Appendix B Appendix to Chapter 3		181
B.1	Proofs	181
B.1.1	Proof of Proposition 4	181
B.1.2	Proof of Lemma 6	182
B.1.3	Proof of Lemma 7	184
B.1.4	Proof of Proposition 8	185
B.1.5	Proof of Corollary 9	188
B.1.6	Proof of Corollary 10	188
B.1.7	Proof of Proposition 11	189

List of Tables

1.1	List of Calibrated Parameters	38
1.2	Transmission mechanism of fundamental risk shock $\Delta\sigma_\rho$	49
1.3	Time series moments of macro uncertainty	59

List of Figures

1.1 Household value function on increasingly adapted sparse grids	32
1.2 Employment Transition Rates over the Business Cycle	41
1.3 Distribution of GDP growth in U.S. postwar data and model simulations	43
1.4 Impulse responses of micro uncertainty to $\Delta\sigma_\rho$	47
1.5 Impulse responses of business cycle aggregates to $\Delta\sigma_\rho$	48
1.6 Impulse responses of business cycle aggregates to $\Delta\sigma_\rho$	52
1.7 Impulse response of endogenous macro uncertainty to $\Delta\rho$	54
1.8 GE interaction between micro, macro uncertainty and aggregate demand	55
1.9 Impulse response of endogenous macro uncertainty to $\Delta\sigma_\rho$	57
1.10 State space representation of Uncertainty Multiplier for different ρ	58
1.11 Anticipation effects near cusp of ZLB	62
2.1 Elementary grids	74
2.2 Union of elementary grids	75
2.3 Nodal basis representation of function	77
2.4 Generalized nodal basis functions	80
2.5 Nodal and hierarchical basis functions for different grids	81
2.6 Function representation in nodal and hierarchical basis	82
2.7 Function representation on adapted sparse grid in the hierarchical basis	84
2.8 Den Haan (2010) metric for solution of the Krusell-Smith model	101
2.9 Error rates for quality ladder model	104

Acknowledgments

I am indebted to Emmanuel Farhi, Xavier Gabaix, Matteo Maggiori and Ludwig Straub for their mentorship, guidance, and support throughout my Ph.D. I am also grateful to Isaiah Andrews, Martin Aragonese, John Campbell, Gabriel Chodorow-Reich, Christopher Clayton, Edward Glaeser, Sam Hanson, Xavier Jaravel, Moritz Lenel, Alisdair McKay, Benjamin Moll, Chris Rycroft, Andrei Shleifer, Adi Sunderam, Jeremy Stein, Paul Tucker, Stephen Terry, Gianluca Violante, Samuel Young, Allen Zhang, as well as seminar participants at Harvard University and HBS for many helpful conversations and comments.

To my mother Ingrid.

Your sacrifice and support made all this possible. Thank you.

Chapter 1

Micro and Macro Uncertainty

1.1 Introduction

Uncertainty at both the micro and macro level rises sharply during economic downturns. At the micro level, households face elevated unemployment risk during crises (e.g. Storesletten *et al.* (2004), Shimer (2005) and Guvenen *et al.* (2014)), and firms experience more volatile sales, cash-flow and productivity growth (e.g. Kehrig (2015) and Bloom *et al.* (2018)). At the macro level, recessions are associated with higher stock market volatility and increased variability in GDP growth (e.g. Bloom (2014)). Much previous work has studied the implications of uncertainty at either the micro or the macro level. A framework that allows for the joint determination of micro and macro uncertainty has thus far remained elusive, in part because it poses considerable methodological challenges. Modeling uncertainty at the micro level requires cross-sectional heterogeneity, while uncertainty at the macro level presupposes aggregate risk.¹

In this paper, I make two main contributions. First, I show that accounting for the interaction between micro and macro uncertainty is crucial to characterize the role uncertainty plays in business cycle fluctuations. The dominant transmission channel of macro uncertainty is its interaction with micro uncertainty. Indeed, in general equilibrium (GE) a strong feedback loop emerges between micro and macro uncertainty. As a result, recessions are associated with

¹Bloom *et al.* (2018) develop a model with cross-sectional firm heterogeneity and study exogenous variation in the micro and macro uncertainty that firms face. My focus, on the other hand, is on the endogenous interaction between uncertainty at the micro and macro level, which is not modeled in their paper.

large endogenous spikes in uncertainty. My second contribution is methodological. I develop a new global solution method for heterogeneous-agent macro models. Leveraging its power, I globally solve a Heterogeneous Agent New Keynesian (“HANK”) model with aggregate risk, counter-cyclical unemployment risk, and a zero lower bound (ZLB) constraint on monetary policy. While a rapidly growing body of work has developed first-order perturbation methods to solve heterogeneous-agent macro models with aggregate risk (see Reiter (2009), Winberry (2020), Ahn *et al.* (2017), Boppart *et al.* (2018), and Auclert *et al.* (2019)), progress on global solution methods has been slower.²

I start my discussion in Section 1.2 by studying the transmission mechanism of macro uncertainty in an illustrative two-period model. Households solve a consumption-savings problem, facing both aggregate wage risk and idiosyncratic unemployment risk. At the heart of my analysis is the empirically motivated assumption that a household’s employment transition probabilities vary with aggregate economic activity. Indeed, the job separation (finding) rate in the data is strongly counter-cyclical (pro-cyclical): Recessions are times of heightened micro uncertainty over households’ job prospects (e.g. Storesletten *et al.* (2004) and Shimer (2012)). The transmission of macro uncertainty in this setting is governed by a set of direct (partial equilibrium) and indirect (general equilibrium) channels, borrowing the language from Kaplan *et al.* (2018). The direct effects of uncertainty are those that arise from a change in households’ beliefs about future shock realizations holding prices constant. Indirect effects arise in general equilibrium as prices respond to the change in household behavior induced by direct effects.

When there is no interaction with micro uncertainty, the only direct effect of macro uncertainty on household behavior to second order is the standard Kimball (1990) precautionary savings motive. A mean-zero spread in next period’s aggregate wage rate exposes employed households to symmetric risk in labor income. Since this effect is marginal, its relevance to households is captured by differentiating marginal utility and, in particular, is proportional to $u'''(\cdot)$. Unlike aggregate wage risk, unemployment risk at the micro level takes the form of discrete and asymmetric jumps. Its relevance is not captured by a derivative of marginal utility

²Variants of the original Krusell and Smith (1998) algorithm largely remain the method of choice. Notable exceptions include Fernández-Villaverde *et al.* (2019) and Pröhl (2019).

but, rather, by the jump scaling factor $u'(c^E) - u'(c^U)$, where $u'(c^E)$ and $u'(c^U)$ are marginal utility conditional on employment and unemployment, respectively. It is in this sense that the prospect of job loss is akin to an idiosyncratic rare disaster from the perspective of households.

When we account for its interaction with unemployment risk, macro uncertainty operates through a set of novel channels. Direct interaction effects emerge because a mean-zero spread in the aggregate shock now also implies increased dispersion in employment outcomes. In general equilibrium, an indirect interaction effect emerges as well: The direct effects of micro and macro uncertainty on household behavior lead to an initial fall in consumer spending and aggregate demand. As a result, the demand for labor falls, which consequently raises (lowers) the job separation (finding) rate. For an employed household, an increase in the job separation rate both lowers expected earnings and increases their variance. The former elicits a consumption-smoothing response and the latter a precautionary savings motive. These novel interaction effects tend to increase households' desired savings in response to an increase in macro uncertainty.

I show that these direct and indirect interaction effects are proportional to the jump scaling factor $u'(c^E) - u'(c^U)$ that is characteristic of uncertainty at the micro level. Households in this setting respond to uncertainty at the macro level not so much because it implies a mean-zero spread in aggregate wages but rather because it translates into disaster risk at the micro level. These results suggest qualitatively that accounting for an interaction with micro uncertainty can substantially alter the transmission mechanism of macro uncertainty.

To evaluate these implications quantitatively, I build a business cycle model in which households face uninsurable unemployment risk in the tradition of Huggett (1993), Aiyagari (1994), McKay *et al.* (2016) and Kaplan *et al.* (2018). I build on the benchmark HANK model of Kaplan *et al.* (2018) along three dimensions. First, I model a zero lower bound (ZLB) constraint on monetary policy, thus explicitly introducing a source of non-linearity at the macro level. Second, I allow for aggregate risk in the form of discount rate shocks.³ Third, and most importantly, I

³Discount rate shocks are a popular proxy for aggregate demand shocks (e.g. Basu and Bundick (2017) and Auclert *et al.* (2020)). I have also solved variants of this model with alternative demand and supply (TFP) shocks. In the presence of nominal rigidities, TFP shocks tend to generate counter-cyclical price inflation (e.g. Gali (2015)). Since the source of macro non-linearity in this model is the ZLB constraint on monetary policy, it is important that inflation is pro-cyclical, which is the case in the presence of aggregate discount rate shocks.

account for the cyclicalities in the job finding and separation rates so that the unemployment risk faced by households varies over the business cycle. I estimate the sensitivity of employment transition rates to changes in economic activity in Current Population Survey (CPS) micro data. This approach follows a long tradition of constructing gross flow data for employment transitions from micro data.⁴ By using the resulting estimates directly in the model, I take as given the link between employment transitions and economic activity.⁵ The sense in which my model generates an endogenous link between micro and macro uncertainty is through economic activity: heightened macro uncertainty depresses aggregate demand which in turn raises unemployment risk. At the micro level, therefore, households face uncertainty over future job prospects. At the macro level, households are exposed to uncertainty over asset prices, interest rates and the aggregate determinants of disposable income, such as the wage rate, transfers, and unemployment insurance (UI) payments. The crucial feature of my model is that uncertainty at the micro level is high precisely at the same time as uncertainty at the macro level.

My first main result is that the transmission mechanism of macro uncertainty in the quantitative model differs starkly from that in a Representative Agent New Keynesian (“RANK”) benchmark.⁶ In the RANK model, the direct precautionary savings response to heightened aggregate uncertainty contributes the largest share to the overall effect on consumption. Using a conservative relative risk aversion coefficient of $\gamma = 2$ in my calibration, this direct precautionary effect is unsurprisingly small. Therefore, the overall effect of macro uncertainty on economic

⁴See for example Marston *et al.* (1976), Abowd and Zellner (1985), Darby *et al.* (1985), Darby *et al.* (1986), Poterba and Summers (1986), Blanchard *et al.* (1990), Shimer (2005), Fujita and Ramey (2009), Elsby *et al.* (2009), and Shimer (2012).

⁵These empirical estimates based on U.S. micro data are intended as a reduced-form summary of the state dependence in employment transition rates that would result from a search-and-matching micro-foundation. Explicitly implementing a search-and-matching block is beyond the scope of this paper and left for future work.

⁶The focus of my paper is on macroeconomic uncertainty that arises endogenously. However, isolating and decomposing the effects of endogenous uncertainty is challenging in a setting with aggregate risk. To be able to study the transmission mechanism of macro uncertainty, I consider an exogenous shock to fundamental risk (similar to Bloom (2009), Basu and Bundick (2017), Bloom *et al.* (2018) and Bayer *et al.* (2019)). The assumption implicit in this strategy is that the transmission mechanism of an exogenous increase in the volatility of discount rate shocks is sufficiently similar to that of an endogenous increase in the volatility of economic activity.

activity in the RANK baseline is quite modest.⁷

In the quantitative model, this direct effect of macro uncertainty is relatively muted. Instead, the interaction with micro uncertainty emerges as the dominant driver of transmission. The quantitative analysis therefore corroborates the importance of the interaction effects identified in the two-period model. Household behavior is more responsive to uncertainty at the macro level precisely because it translates into unemployment risk. Similarly, and in the spirit of Kaplan *et al.* (2018), other indirect channels working through portfolio returns and the aggregate determinants of disposable income become important relative to the direct precautionary channel. Therefore, the overall effect of macro uncertainty on economic activity is large in my model precisely because it works through micro uncertainty and other indirect channels. I show that the peak response of output to an increase in macro uncertainty is 5 to 8 times larger than in the associated RANK benchmark. The interaction between micro and macro uncertainty is the main source of this amplification.

An overarching theme of my analysis is that the behavior of uncertainty changes drastically during economic crises. I show that the peak decline in output in response to a given increase in macro uncertainty is 50% larger when the economy is already at the cusp of the ZLB than during normal times. Close to the ZLB, the relationship between economic activity and aggregate risk exhibits a degree of negative skewness. As a result, a mean-zero spread in aggregate risk leads to a contraction of economic activity in expectation. The implications of such skewness at the macro level are not unlike the skewness households face at the micro level. Indeed, I show that non-linearity at the macro level strongly interacts with non-linearity at the micro level: The relative importance of its interaction with micro uncertainty in the transmission of macro uncertainty further rises during economic crises. Identifying the importance of the ZLB crisis region for the behavior of uncertainty in my model is only possible because I use a global solution method.⁸

⁷This is consistent with the results in Basu and Bundick (2017). They show that a fundamental risk shock induces co-movement in output, consumption, investment and hours, but they impart households with Epstein-Zin preferences and a relative risk aversion coefficient of 80 to generate quantitatively meaningful responses.

⁸Overall, these results highlight that uncertainty can have large effects on consumption even in a setting where household behavior is not perfectly forward-looking. It is well known that the effective planning horizon of households in an incomplete markets setting is shortened in the presence of borrowing constraints (e.g. McKay

My second main result is that the interaction between micro and macro uncertainty has implications not only for the transmission mechanism of macro uncertainty but also for its endogenous responsiveness to changes in economic activity. I show that macro uncertainty in my model rises endogenously during economic downturns. That is, recessions are accompanied by endogenous spikes in uncertainty even in the absence of exogenous second-moment shocks. Crucially, the sensitivity of macro uncertainty to economic activity is dampened substantially when I shut off its interaction with micro uncertainty. Concretely, I show that endogenous macro uncertainty is 4 times more responsive to a negative, first-moment discount rate shock when I account for the interaction with micro uncertainty. Indeed, when I hold unemployment risk constant, macro uncertainty hardly responds at all to discount rate shocks during normal times.

Macroeconomic uncertainty responds endogenously to changes in economic activity through two main channels in my model. The first channel centers around the zero lower bound constraint. When the economy is at the ZLB, monetary policy can no longer accommodate negative demand shocks, whose effects on economic activity are consequently amplified. Even when the nominal interest rate is still positive, expansionary monetary policy moves the economy closer to the ZLB, thus raising the likelihood that policy will be constrained in the future. A given increase in macro uncertainty resulting from proximity to the ZLB has a larger effect on aggregate demand when its transmission works through micro uncertainty, thus pushing the economy even closer to the ZLB. To capture this economic force, it is crucial to employ a global solution method.

The second channel results from the counter-cyclical behavior in households' average marginal propensity to consume (MPC). As economic activity contracts, households become unemployed and draw down their liquid cash buffers, thus moving closer to their borrowing constraints at the micro level. The prevalence of "hand-to-mouth" behavior grows, which implies an increase in the average household's MPC. Consumer spending, and by implication aggregate activity, thus become more sensitive to further demand shocks, which represents an increase in macro

et al. (2016) and Kaplan *et al.* (2018)). Households directly at the borrowing constraint consume their disposable income "hand-to-mouth". But even those households that are close to but not at the constraint only try to smooth consumption until they expect to reach the constraint. Unlike the representative household in the RANK benchmark, most households in my model are close to or at their borrowing constraint and have relatively short planning horizons.

uncertainty. By contrast, the counter-cyclical of the average MPC is considerably dampened when employment transition rates are held constant over the business cycle. The standard Krusell and Smith (1998) algorithm struggles to account for these shifts in the household distribution. The global solution method developed in this paper therefore plays a key role in my ability to study endogenous uncertainty spikes in this model.

In general equilibrium, therefore, a strong feedback loop emerges between uncertainty and economic activity: When activity contracts, uncertainty about the future rises, which itself depresses aggregate demand further. This feedback loop can be instructively characterized as an “Uncertainty Multiplier”, which measures how much endogenous amplification there is in macro uncertainty.⁹ I show that this Uncertainty Multiplier is high precisely when we account for the cyclical of unemployment risk, and when the economy is already in a recession. Indeed, this feedback loop between uncertainty and activity allows my model to match the time series moments of various macro uncertainty proxies in the data without requiring exogenous second-moment shocks: Macro uncertainty in the model is strongly counter-cyclical, highly persistent and exhibits large positive skewness and kurtosis. Overall, uncertainty emerges as both a driver and a byproduct of business cycle fluctuations. Accounting for the interaction between micro and macro uncertainty is thus crucial to understand the broader role that uncertainty plays in macroeconomic fluctuations.

The interaction between micro and macro uncertainty has far-reaching implications for central questions in business cycle analysis, several of which I discuss in Sections 1.7 and 1.8, and in the appendix. First, ZLB spells become more frequent and more persistent. I also show that the interaction between the ZLB and a paradox of thrift dynamic, by which households increase savings in anticipation of reaching the ZLB, is amplified. Second, an interplay between uncertainty at the micro and macro level provides a new perspective on the welfare cost of business cycles. I show that the implied share of consumption households are willing to forego to instead “live” in a representative-agent economy is 3.9%, or about two orders of magnitude

⁹Similar uncertainty multipliers, operative at either the micro or macro level, have been discussed in other settings. For example, see Plante *et al.* (2018) for an application with zero lower bound constraint, Ravn and Sterk (2017) for an example with counter-cyclical unemployment risk, and Fernández-Villaverde *et al.* (2019) for a setting with financial constraints.

larger than the original estimates in Lucas (1987) and Lucas (2003). In the appendix, I study a series of policy experiments and show that stabilization policy in my setting operates through a novel set of micro and macro uncertainty channels.

Methodological Contribution. The methodological contribution of my paper is a new global solution method for heterogeneous-agent macro models with aggregate risk. I show that this new method is key to solve the quantitative model with time-varying unemployment risk and occasionally-binding ZLB constraint, and in turn study the interaction between micro and macro uncertainty. The main challenge in numerically solving my model is that the entire cross-sectional distribution of agents, an infinite-dimensional object, becomes part of the aggregate state space. Let \mathbf{x}_t denote the vector of idiosyncratic state variables and $g_t(\mathbf{x})$ the cross-sectional distribution of agents. In this paper, I work with finite-dimensional distribution approximations of the form

$$F(\alpha_t)(\mathbf{x}) \approx g_t(\mathbf{x}).$$

For illustration, it is easiest to think of F as a set of basis functions that are parameterized by the time-varying $\alpha_t \in \mathbb{R}^N$. While representations of the form $F(\alpha_t)(\mathbf{x}) \approx g_t(\mathbf{x})$ are commonly used in the context of local perturbation methods, they have proven intractable in the context of global methods due to the curse of dimensionality.¹⁰ I make three contributions that help overcome this challenge:

1. Most global methods currently in use, such as the seminal Krusell and Smith (1998) algorithm, work with a finite set of moments to approximate the distribution of agents. The costliest step of these algorithms is finding an internally consistent law of motion for these moments. I show in Section 1.4.3 that, for a large class of models, the coefficients α_t follow a diffusion process with drift $\mu_{\alpha,t}$ and volatility $\sigma_{\alpha,t}$, and I derive analytical formulas for these objects that can be easily computed. In this sense, and in sharp contrast to the Krusell and Smith (1998) algorithm, finding the consistent law of motion incurs almost no

¹⁰My paper builds on the important contribution of Winberry (2020) who uses a distribution representation of this form in the context of a local perturbation method.

increase in numerical complexity.¹¹

2. While $F(\cdot)$ can be chosen from a parametric family, I develop a non-parametric algorithm in Section 1.4.4 that delivers substantial efficiency gains especially when the idiosyncratic state space of agents is high-dimensional.
3. For most economic applications of interest, accurate approximations of the distribution will require high-dimensional $F(\alpha_t)(x)$. Global methods will therefore quickly encounter the curse of dimensionality. In Schaab and Zhang (2020), we develop an adaptive sparse grid library for solving partial differential equations that can overcome the curse of dimensionality in high dimensions. Using this library and leveraging the results developed in this paper, I can solve the benchmark Krusell and Smith (1998) model with a $F(\alpha_t)(x)$ representation in over 20 dimensions, that is $\alpha_t \in \mathbb{R}^{20}$.

Literature Review. This paper is most directly related to a long literature studying the role of uncertainty in business cycle fluctuations. One prominent strand of this literature, of which Bernanke (1983) is an early example, follows the seminal contribution of Bloom (2009) and asks whether uncertainty can drive business cycles. This group of papers considers the implications of an exogenous increase in micro or macro uncertainty, prompting a precautionary savings response among households (Leduc and Liu (2016), Basu and Bundick (2017), Bayer *et al.* (2019)), a wait-and-see response by firms (Bloom (2009), Bloom *et al.* (2018)) or a tightening of financial constraints (Gilchrist *et al.* (2014), Arellano *et al.* (2019)).¹² An alternative approach to study the effects of uncertainty on economic activity uses vector autoregression (VAR) estimates (see for example Bloom (2009), Ludvigson *et al.* (2015) or Basu and Bundick (2017)). Relative to this literature, I show that accounting for the interaction between micro and macro uncertainty qualitatively changes the transmission mechanism of an uncertainty shock, substantially amplifying

¹¹I build on the closely related contribution of Ahn *et al.* (2017) and generalize this state space reduction technique to settings where the cross-sectional distribution of agents itself is stochastic. Asset pricing models with portfolio choice problems are typically of this kind.

¹²See Fernández-Villaverde *et al.* (2015b) for a quantitative analysis of a policy uncertainty shock.

its effect on activity.¹³

A largely distinct strand of literature argues that uncertainty arises endogenously as a byproduct of economic crises.¹⁴ Several channels have been proposed through which a contraction in economic activity may spur uncertainty: When economic activity falls, economic agents interact less frequently, stifling the spread of information (Van Nieuwerburgh and Veldkamp (2006), Fajgelbaum *et al.* (2017), Straub and Ulbricht (2017)); policymakers may resort to adopting untested policies, raising uncertainty (Pástor and Veronesi (2013)); firms may take riskier and more experimental actions during bad times (Bachmann *et al.* (2011)). The endogenous responsiveness of uncertainty to economic activity in my model works largely through two channels: the ZLB, and counter-cyclical MPCs. Plante *et al.* (2018) similarly make the argument that, during bad times as the economy approaches the ZLB, policy becomes further incapacitated, which raises uncertainty.¹⁵ My paper is also closely related to Fernández-Villaverde *et al.* (2019). They study endogenous uncertainty in a heterogeneous-household setting in the tradition of Huggett (1993) that also features an aggregate financial friction.¹⁶

My focus on time-varying micro uncertainty in the form of counter-cyclical unemployment risk is shared by a large body of work that documents the cyclicity of earnings risk and employment transitions in the data (e.g. Storesletten *et al.* (2004), Shimer (2005) and Guvenen *et al.* (2014)), and studies its implications analytically and quantitatively (e.g. Ravn and Sterk (2016), Schmidt (2016), McKay (2017), Acharya and Dogra (2020)). In this paper, I study the interaction between unemployment risk and macro uncertainty: Households respond to an increase in macro uncertainty in large part because it translates into micro uncertainty. Patterson (2019) documents systematic heterogeneity in household exposure to cyclical earnings risk. Taking into account heterogeneous incidence is left for future work. My paper is also particularly

¹³There is also a large literature that tries to measure uncertainty in the data. See Bloom (2014) for an overview.

¹⁴A smaller body of work seeks to determine whether the counter-cyclicity of uncertainty is a result of exogenous shocks or rather an endogenous response. See for example Ludvigson *et al.* (2015) and Berger and Vavra (2019).

¹⁵While they make this point in a representative-agent New Keynesian model, I show that accounting for cross-sectional heterogeneity and, in particular, the interaction between micro and macro uncertainty is crucial.

¹⁶Relative to their paper, my focus is on nominal rigidities and the ZLB constraint. Households in my setting can trade two assets. Lastly, an endogenous interaction between micro and macro uncertainty is at the heart of my mechanism whereas the earnings risk households face in their model is constant.

closely related to Krusell *et al.* (2010) and Den Haan *et al.* (2018). Relative to their work, my main focus is on the transmission mechanism of fundamental risk or macro uncertainty shocks. The ZLB also plays a larger role in my paper, generating significant non-linearity at the macro level. In particular, my paper emphasizes the interplay between micro and macro uncertainty which both rise endogenously during economic downturns.

My paper also adds to the burgeoning heterogeneous-agent New Keynesian (HANK) literature.¹⁷ This is the first paper of this class that studies macroeconomic uncertainty and features an occasionally binding macro constraint; both of these features require a higher-order or indeed global solution method.¹⁸ My main contribution to this literature is to show that the transmission mechanism of uncertainty changes qualitatively when we account for cross-sectional household heterogeneity: the effect through micro uncertainty becomes the new dominant transmission channel of a macro uncertainty shock.

Finally, I build on an extensive body of work that has developed global solution methods for heterogeneous-agent macro models.¹⁹ My paper also builds on the important contributions of Winberry (2020) and Ahn *et al.* (2017) who propose a similar finite-dimensional distribution representation as I do in the context of local perturbation methods.

1.2 Illustrative Two-Period Example

The goal of this section is to develop intuition for the economic mechanism driving my results: the interaction between micro and macro uncertainty and its implications for household behavior. I characterize the effect of uncertainty on household consumption in a stylized two-period model of consumption and savings decisions. Households in this setting face unemployment risk at the

¹⁷See Oh and Reis (2012), Lorenzoni (2008), McKay and Reis (2016), McKay *et al.* (2016), Werning (2015), Challe *et al.* (2017), Kaplan *et al.* (2018), Auclert *et al.* (2019), Auclert *et al.* (2018), Auclert *et al.* (2020), Bayer *et al.* (2019), Ottonello and Winberry (2017), Acharya and Dogra (2020), and Bilbiie (2020). This list is non-exhaustive.

¹⁸There is a large literature that studies the zero lower bound constraint quantitatively in representative-agent settings. See for example Christiano *et al.* (2011), Fernández-Villaverde *et al.* (2015a), Nakata (2017) and Plante *et al.* (2018).

¹⁹For example, see Den Haan (1996), Den Haan and Others (1997), Krusell and Smith (1998), Reiter (2010), Algan *et al.* (2008), Algan *et al.* (2014), Brunnermeier and Sannikov (2014), Brumm and Scheidegger (2017), Duarte (2018), Pröhl (2019), and Fernández-Villaverde *et al.* (2019).

micro level and uncertainty over wage growth at the macro level. The critical assumption I make is that households' employment transition probabilities are a function of aggregate economic activity.

Setting. There are two periods, $t = 0, 1$. Aggregate risk is represented by the normal random variable $\sigma\epsilon \sim \sigma\mathcal{N}(0, 1)$, which is realized at the beginning of period 1. While I leave the macro block of the model largely unspecified, I assume and work with a notion of aggregate economic activity which I denote by Y_t . The only structure I have to impose is that economic activity in period 1 responds to the realization of aggregate risk, that is $Y_1 = Y_1(\sigma\epsilon)$.²⁰

Households. A continuum of households i make consumption and savings decisions, facing uncertainty at both the micro and macro level. Household i 's budget constraints are given by

$$c_{i,0} + a_{i,1} = a_{i,0} + y_{i,0}z_{i,0}$$

$$c_{i,1} = Ra_{i,1} + y_{i,1}z_{i,1}.$$

In period 0, households consume, $c_{i,0}$, and save, $a_{i,1}$, out of initial wealth, $a_{i,0}$, and total labor income given by $y_{i,0}z_{i,0}$. In period 1, household consumption is equal to the gross return on savings and labor income. Labor supply is inelastic.

The key object in this stylized setting is household labor income, which comprises a component that loads on the aggregate state, $y_{i,t}$, and a purely idiosyncratic term, $z_{i,t}$. The idiosyncratic component corresponds to the household's employment status, with $z_{i,t} \in \{0, 1\}$. Conditional on employment, $z_{i,t} = 1$, $y_{i,t}$ can be thought of as a wage. In particular, I assume that a household's wage is directly proportional to economic activity, $y_{i,t} = \gamma_i Y_t$, as in Werning (2015).

The main assumption I make, which introduces a meaningful interaction between micro

²⁰More formally, the discussion in this section is valid as long as the general equilibrium block satisfies the following restriction. Let X_t denote the vector of all macroeconomic aggregates in period t . Then there are sets of equations, which I dub the macro block, $\mathcal{H}_0(X_0, \mathbb{E}_0(X_1)) = 0$ and $\mathcal{H}_1(X_0, X_1, \sigma\epsilon) = 0$. In particular, the assumption that only the first moment of future aggregate states of the economy affects the allocation in period 0 guarantees that macro uncertainty has no effect on household behavior to first order. In this setting, we have $Y_1 = Y_1(X_0, \sigma\epsilon)$. In response to an increase in σ , there is a direct (partial equilibrium) effect on Y_1 through its second argument and an indirect (general equilibrium) effect through its first argument.

and macro uncertainty, is that the unemployment risk faced by households is counter-cyclical. In particular, I assume that the probability that household i becomes or remains employed in period 1 directly depends on economic activity, with

$$\mathbb{P}(z_{i,1} = 1 \mid z_{i,0}, \epsilon) = p_i(Y_1),$$

where p_i corresponds to the job finding rate when $z_{i,0} = 0$ and one minus the job separation rate when $z_{i,0} = 1$.

Household preferences are defined over consumption, given by

$$\mathbb{E}_0 \sum_{t=0}^1 \beta^t U(c_{i,t}).$$

The household's consumption and savings decision is then characterized by a standard Euler equation. Namely,

$$U'(c_{i,0}) = \beta R \mathbb{E}_0 \left[U'(c_{i,1}^u) (1 - p_i(Y_1)) \right] + \beta R \mathbb{E}_0 \left[U'(c_{i,1}^e) p_i(Y_1) \right]$$

where I have used the law of total probability to split expected marginal utility in period 1 into an unemployment and an employment term. $U'(c_{i,1}^u)$ is the marginal utility of consumption conditional on unemployment, and it is multiplied by the probability of unemployment given economic activity Y_1 . Similarly, $U'(c_{i,1}^e)$ is the marginal utility of consumption conditional on employment, and it is multiplied by the probability of employment.

To think about the effects of uncertainty on household behavior, I consider a comparative static in σ , which is analogous to a macro uncertainty shock in this illustrative model. The following result characterizes how household consumption depends on σ locally around $\sigma \approx 0$. That is, I perform a Taylor expansion around an economy that features micro uncertainty in the form of employment transitions but no aggregate risk.

Proposition 1. *To second order,*

$$c_{i,0}(\sigma) \approx c_{i,0}(0) + \frac{1}{2} \frac{d^2 c_{i,0}}{d\sigma^2}(0) \sigma^2,$$

where

$$\begin{aligned}
\frac{d^2 c_{i,0}}{d\sigma^2} = & \underbrace{\left\{ \frac{U'''(c_{i,1}^e) \gamma_i^2}{U''(c_{i,0})} p_i \mathbb{E}_0 \left[\left(\frac{\partial Y_1}{\partial \sigma} \right)^2 \right] \right\}}_{\text{① Pure macro (wage) uncertainty}} + \underbrace{\frac{U'(c_{i,1}^e) - U'(c_{i,1}^u)}{U''(c_{i,0})}}_{\text{Micro uncertainty}} \left(\underbrace{p_i'' \mathbb{E}_0 \left[\left(\frac{\partial Y_1}{\partial \sigma} \right)^2 \right]}_{\text{② Direct (PE) effect}} + \underbrace{p_i' \mathbb{E}_0 \left[\left(\frac{d^2 Y_1}{d\sigma^2} \right) \right]}_{\text{③ Indirect (GE) effect}} \right) \\
& + \underbrace{2 \frac{U''(c_{i,1}^e) \gamma_i}{U''(c_{i,0})} p_i' \mathbb{E}_0 \left[\left(\frac{\partial Y_1}{\partial \sigma} \right)^2 \right]}_{\text{④ Cov micro} \times \text{macro risk}} \} \times \beta R \text{MPS}_{i,0} + \underbrace{\Delta_i}_{\substack{\text{"Standard" effects} \\ \text{(no interaction with uncertainty)}}}
\end{aligned}$$

All objects are evaluated in the limit as $\sigma \rightarrow 0$, and $\text{MPS}_{i,0}$ is household i 's marginal propensity to save in period 0.

Macro uncertainty affects household consumption in this model through four channels:

1. As in any representative-agent model, aggregate risk elicits the standard precautionary savings motive that is proportional to $U''' > 0$ (Kimball (1990)), channel ①. Only employed households are exposed to wage risk, so this term is scaled by the probability of employment, p_i . Their exposure is proportional to the sensitivity of wages to the aggregate shock ϵ , which, to second order, is given by $(\gamma_i Y_1')^2$.

The remaining three channels characterize the effect of macro uncertainty on household consumption *through* its interaction with micro uncertainty.

Aggregate wage risk exposes households to marginal changes in consumption, whose relevance is captured by differentiating marginal utility. Micro uncertainty in the form of employment transitions, on the other hand, represents an *idiosyncratic rare disaster* for households: it leads to discrete jumps in consumption and, therefore, marginal utility. Its relevance is not captured by a derivative of marginal utility but, rather, by a *jump* scaling factor. The effect on marginal utility of a transition from unemployment to employment is thus given by $U'(c_{i,1}^e) - U'(c_{i,1}^u)$. It is in precisely this sense that the prospect of job loss is akin to a rare disaster, and consequently much more prominent, from the perspective of households.

2. Both channels ② and ③ represent the effect of macro uncertainty on micro uncertainty, and are thus scaled by $U'(c_{i,1}^e) - U'(c_{i,1}^u) < 0$. Channel ② corresponds to the *direct* or

partial equilibrium effect.²¹ This channel is operative only when $p_i'' \neq 0$, so that there is a non-linearity at the micro level in the response of employment transition rates to economic activity. I present some empirical evidence for such non-linearities in Section 1.5.2. Consider the case of an employed household. Intuitively, if the job *separation* rate is strictly convex as a function of economic activity, $p_i'' < 0$, then a mean-zero spread in aggregate risk leads to an increase in the household's expected probability of unemployment. Through this interaction with unemployment risk, macro uncertainty has a first-moment effect on expected earnings and induces desired savings.

3. Macro uncertainty has an effect on micro uncertainty to second order, even if the job finding and separation rates are only linear, $p_i' \neq 0$. Channel ③ represents this *indirect* or general equilibrium effect. An increase in uncertainty will, to second order, elicit precautionary savings (e.g. channel ①) and thus lead to a contraction in aggregate economic activity. If this contraction in activity is persistent and propagates into period 1, then households expect an associated GE effect on employment transition rates.²² Simply put, if a macro uncertainty shock leads to a persistent recession, then households expect elevated unemployment risk going forward.²³ The indirect GE effect of channel ③ is the dominant transmission channel of macro uncertainty in the quantitative model: households respond to an increase in macro uncertainty precisely because it translates into

²¹I use “direct” and “indirect” in the sense of Kaplan *et al.* (2018).

²²In the continuous-time quantitative model of Section 1.3, this GE effect will be *contemporaneous*. Therefore, an indepth discussion of the sources of persistence and propagation is unnecessary at this point.

²³When the economy's general equilibrium block takes the form discussed in the previous footnote, then this GE effect can formally be further decomposed into two components. When $Y_1 = Y_1(X_0, \sigma\epsilon)$, we have

$$\mathbb{E}_0 \left[\frac{d^2 Y_1}{d\sigma^2} \right] = \underbrace{\mathbb{E}_0 \left[\frac{\partial Y_1}{\partial X_0} \right] \frac{d^2 X_0}{d\sigma^2}}_{\text{Endogenous GE effect}} + \underbrace{\mathbb{E}_0 \left[\frac{\partial^2 Y_1}{\partial \sigma^2} \right]}_{\text{Macro non-linearity}} .$$

The first term highlights that, to second order, a macro uncertainty shock adversely affects aggregates in period 0, $d^2 X_0 / d\sigma^2$. This can lead to a persistent recession in period 1, $\mathbb{E}_0[\partial Y_1 / \partial X_0]$, and raise the expected probability of job loss. The second term emphasizes that, even holding the other macro aggregates X_0 fixed, there will be a similar effect if there is skewness at the aggregate level. When the economy is close to or in a crisis region, then a mean-zero spread in aggregate risk will have a first-moment impact on activity because, due to negative skewness, a negative shock is amplified more than a similarly sized positive shock. The large negative skewness exhibited by U.S. business cycles speaks to the relevance of this economic force.

micro uncertainty.

4. Finally, there is a covariance between risks at the micro and macro level. Formally, channel ④ represents the effect of σ , to second order, on the covariance between marginal utility and employment transition rates. When the probability of transitioning into employment state j is high for precisely those realizations of ϵ that also imply a high marginal utility of employment state j , then this channel represents another source of risk for households and induces precautionary savings. If the covariance structure is reversed, then this represents a hedging term.

The behavioral response in consumption is also subject to other effects, specifically a set of interest rate and earnings effects, that I suppress here. These effects are “standard” in the sense that they would also emerge in a representative-agent setting and do not interact with uncertainty in meaningful ways.

1.3 A HANK Model with Micro and Macro Uncertainty

The quantitative model takes as its starting point the heterogeneous-household model proposed by Kaplan *et al.* (2018). Households face uninsurable earnings risk but can trade liquid and illiquid assets under incomplete markets. I depart from this benchmark in three important ways. First, I model aggregate uncertainty in the form of discount rate shocks. Second as in the illustrative model in Section 1.2, I assume that unemployment risk varies over the business cycle. Third, I introduce a zero lower bound (ZLB) constraint on monetary policy.

1.3.1 Households

The economy is populated by a continuum of households. Facing both idiosyncratic unemployment risk and aggregate uncertainty, households make consumption, savings and portfolio allocation decisions across time. The idiosyncratic state of a household consists of its portfolio position, made up of liquid assets a_t and illiquid assets k_t , and its employment status z_t .

Household preferences are defined over consumption and labor, given by

$$\max \mathbb{E}_0 \int_0^\infty e^{-\int_0^t (\rho_s + \zeta) ds} u(c_t, h_t) dt, \quad (1.1)$$

where c_t is the rate of consumption and h_t denotes the rate at which work hours are supplied. As in the canonical New Keynesian model, c_t is a consumption basket which is itself comprised of intermediate goods. This final good basket is priced at the consumer price index (CPI) P_t . Households die at rate ζ . At the same time, an equal mass of new households is formed with zero liquid and illiquid assets.

Discount rate shocks. The household's effective discount rate is given by $\rho_t + \zeta$. All households share a time-varying discount rate, ρ_t , which is the source of aggregate demand shocks in my model.²⁴ It follows a continuous-time AR(1) process, given by

$$d\rho_t = \theta_\rho (\bar{\rho} - \rho_t) dt + \sigma_\rho dB_t,$$

where B_t is a standard Brownian motion.²⁵ Discount rate shocks are a popular proxy for aggregate demand shocks.

Budget constraint. Households trade in two asset markets, one for bonds (liquid assets) and one for capital (illiquid assets). I denote the household's liquid asset position by a_t and its illiquid asset position by k_t . Households are endowed with an investment technology that transforms $Q_t \iota_t + P_t \psi(\iota_t, k_t)$ units of the numeraire into ι_t units of capital, where Q_t is the price of capital investment and P_t is the CPI. I denote the real price of capital by $q_t = Q_t / P_t$. ψ_t represents an investment adjustment cost and is the source of capital's illiquidity. Households are thus the direct owners of capital in this model, which they rent to firms in an economy-wide, competitive rental market.

Following Blanchard (1985), I introduce perfect annuity markets, in which households can

²⁴Basu and Bundick (2017), an important reference point for my quantitative results, also build a model with aggregate discount rate shocks. Using the same underlying shock process makes a direct comparison of results easier. Auclert *et al.* (2020) also use a discount rate shock as one of their demand shocks.

²⁵I have solved versions of the model with alternative demand and supply shocks. They are available on request.

trade claims on their remaining wealth at time of death. They pledge this wealth to a risk-neutral insurance company that, in turn, compensates households with a flow annuity payment at a rate ζ times their current asset positions. This is exactly the payment rate that makes the insurance company break even in expectation. See Appendix A.1.4 for details. Introducing household death rates is a commonly used technique to ensure stationarity in the wealth distribution.

A household's liquid asset position evolves according to

$$\dot{a}_t = (r_t + \zeta)a_t + k_t \frac{dR_t}{dt} + e_t - q_t \iota_t - \psi(\iota_t, k_t) - c_t. \quad (1.2)$$

This budget constraint is derived from its nominal analog (see Appendix A.1.1 for details). The real rate of return on the liquid asset consists of the real riskfree rate r_t and the rate of annuity payments ζ . Capital earns a real rate of return dR_t , which I will further specify in Section 1.3.6 after presenting the rest of the model. Wage payments and rebates are collected in the earnings variable e_t , given by

$$e_t = (1 - \tau^{\text{lab}})z_t w_t h_t + \tau_t^{\text{lump}} + \tau^{\text{UI}}(z_t),$$

where $w_t = W_t/P_t$ is the real wage, τ_t^{lump} is a set of lump-sum rebates and $\tau^{\text{UI}}(z_t)$ denotes unemployment insurance payments that explicitly depend on the household's employment status. Finally, households consume and buy illiquid assets at relative price q_t , subject to the adjustment cost ψ .

The household's illiquid asset position evolves as

$$\dot{k}_t = (\zeta - \delta)k_t + \iota_t, \quad (1.3)$$

where δ denotes capital depreciation and ζ the rate of annuity payments. Since households are the direct owners of capital, they also incur depreciation. For convenience, I will later use the shorthand notation s_t and m_t to refer to the drift in the household's liquid and illiquid asset positions, respectively.

Finally, households' portfolio allocation is subject to a borrowing constraint on liquid assets, $a_t \geq \underline{a}$ with $0 > \underline{a}$, and a short-sale constraint on capital, $k_t \geq 0$.

The household's dynamic problem is therefore to maximize (1.1) subject to budget constraints (1.2) and (1.3), as well as the borrowing and short-sale constraints. Households take

as given the laws of motion for their employment status and macroeconomic aggregates. I denote the resulting policy functions in terms of the household's idiosyncratic state variables as $c_t(a, k, z)$, $h_t(a, k, z)$ and $\iota_t(a, k, z)$.

Appendix A.1 provides additional details on the household problem. I derive a recursive representation and the associated consumption Euler equation in Appendix A.1.1. In Appendix A.1.3, I discuss the implications of inflation risk for the household's portfolio equations.

Investment adjustment cost. I adopt the functional form for the household's investment adjustment cost that is used by Kaplan *et al.* (2018). In particular,

$$\psi(\iota_t, k_t) = \psi_0 |\iota_t| + \psi_1 \left(\frac{|\iota_t|}{k_t} \right)^2 k_t.$$

Unemployment risk. Households face uninsurable earnings risk that is encoded in the state variable z_t , which follows a two-state Markov process. These two states are given by $z_t \in \{z^E, z^U\}$ and are thus meant to represent employment and unemployment. In practice, I set $z^E = 1$ and $z^U = 0$ as in the two-period model.

Formally, let N_t be a standard Poisson process and let $j \in \{E, U\}$ index the household's current earnings state. Then the evolution of the earnings state follows

$$dz_t^j = (z^{-j} - z^j) dN_t(\lambda_t^j),$$

where λ_t^j is the Poisson arrival rate. These transition rates are the continuous-time analog of the transition probabilities p_i in the two-period model. For now, I specify λ^E and λ^U as arbitrary functions on the aggregate state space, which the household takes as given. In Section 1.5.2, I estimate the sensitivity of employment transition rates to changes in economic activity in Current Population Survey (CPS) micro data. By using the resulting estimates directly in the model, I take as given the link between employment transitions and economic activity.²⁶ The sense in which my model generates an endogenous link between micro and macro uncertainty

²⁶These empirical estimates are intended as a reduced-form proxy for the state dependence in employment transition rates that would result, for example, from a search-and-matching micro-foundation. Explicitly implementing a search-and-matching block is beyond the scope of this paper and left for future work.

is through changes in economic activity: heightened macro uncertainty depresses aggregate demand which in turn raises unemployment risk.

1.3.2 Firms

The structure of the goods producing sector is as in the standard New Keynesian model. Monopolistic intermediate producers, which I will simply refer to as firms, sell differentiated varieties to a retailer that aggregates these into a composite final consumption good.

Retailer. The aggregation technology of the retailer is given by

$$Y_t = \left(\int_0^1 Y_t(j)^{\frac{e^f-1}{e^f}} dj \right)^{\frac{e^f}{e^f-1}},$$

where $Y_t(j)$ is the output produced by intermediate firm j and e^f denotes the elasticity of substitution across intermediate varieties. The retailer's cost minimization problem gives rise to the standard demand function for intermediate inputs, $Y_t(j) = (P_t(j)/P_t)^{-e^f} Y_t$, where $P_t(j)$ is the price of firm j .

Firms. Intermediate goods producers operate a technology that combines capital and labor. Firm j 's production function is given by

$$Y_t(j) = K_t(j)^{1-\beta} L_t(j)^\beta, \tag{1.4}$$

where β denotes the labor share.

Firm j demands work hours at rate $L_t(j)$, for which it pays the nominal wage rate W_t . There is an economy-wide rental market for capital in which firms rent capital from households at nominal rental rate i_t^k . All firms face the same prices and take them as given. Firm j 's nominal profits are thus given by $\Pi_t(j) = P_t(j)Y_t(j) - W_t L_t(j) - i_t^k K_t(j)$. Standard cost minimization implies a composite nominal marginal cost of

$$MC_t = \frac{1}{\beta^\beta (1-\beta)^{1-\beta}} \left(i_t^k\right)^{1-\beta} \left(W_t\right)^\beta,$$

and I define the real marginal cost as $mc_t = MC_t/P_t$, so that the real rental rate of capital is

analogously given by $r_t^k = i_t^k / P_t$.

Dynamic price setting. While the choice of factor inputs is a static one, firms are monopolistically competitive and face a dynamic pricing problem subject to price adjustment costs. I adopt the quadratic specification of Rotemberg (1982).

Define $\pi_t(j)$ to be the instantaneous rate of inflation chosen by firm j . This rate is chosen to maximize an appropriately discounted sum of all future expected profits subject to an adjustment cost which I specify in terms of firm utility. This dynamic problem is given by

$$\max_{\{\pi_t(j)\}} \mathbb{E}_0 \int_0^\infty e^{-\int_0^t i_s^k ds} \left[(1 - mc_t) P_t(j) Y_t(j) - \Lambda(\pi_t(j)) \right] dt, \quad (1.5)$$

where the effective discount rate is in terms of the cost of capital i_t^k . See Appendix A.1 for additional details on the firm pricing problem.

1.3.3 Capital producer

For expositional clarity, I explicitly specify a capital producing sector.²⁷ The capital producer is owned by households. It is endowed with a technology that transforms the final consumption good into capital. Concretely, it requires $I_t + \Phi(I_t/K_t)K_t$ units of the final good to produce I_t units of capital. The capital producer sells all new capital to households at the nominal price $Q_t = P_t q_t$.

The real rate of profit earned by the capital producer, and paid out to households, is thus given by

$$\Pi_t^Q = q_t I_t - I_t - \Phi\left(\frac{I_t}{K_t}\right) K_t, \quad (1.6)$$

where I_t should be interpreted as the rate at which new capital, i.e. the latest vintage, is generated. The associated optimality condition for static profit maximization is then given by

$$q_t = 1 + \Phi'\left(\frac{I_t}{K_t}\right). \quad (1.7)$$

Finally, accounting for the depreciation of capital incurred by households, the evolution of the

²⁷Capital production could be subsumed in the household problem, as in Brunnermeier and Sannikov (2014) for example, as long as one is careful to distinguish between the idiosyncratic and aggregate adjustment costs.

economy's aggregate capital stock is given by

$$\dot{K}_t = I_t - \delta K_t. \quad (1.8)$$

1.3.4 Nominal wage stickiness

I follow a long tradition in the wage rigidity literature and model household labor supply h_t as determined by union labor demand (see Erceg *et al.* (2000), Schmitt-Grohé and Uribe (2005), and Auclert *et al.* (2020)). There is a continuum of labor unions indexed by $k \in [0, 1]$. Households supply labor to each of these unions, which in turn bundle labor and pass on a differentiated labor variety to an aggregate labor packer. This structure is meant to mimic the analogous structure in the standard price stickiness setup where intermediate producers sell differentiated varieties to a final retailer. The final labor packer aggregates all labor in the economy into a composite labor factor, which is then used by firms in the production of intermediate goods.

Since each household supplies labor to each of the k unions, we have $h_t = \int h_{k,t} dk$. Union k 's aggregation technology combines all households' effective work hours, $z_t h_{k,t}$, into a union-specific labor variety $L_{k,t}$. An aggregate, competitive labor packer combines these intermediate inputs into an aggregate labor supply basket according to

$$L_t = \left(\int L_{k,t}^{\frac{\epsilon^w - 1}{\epsilon^w}} dk \right)^{\frac{\epsilon^w}{\epsilon^w - 1}}.$$

It sells this composite good at the nominal wage rate W_t to intermediate goods producers. The demand function of the labor packer for union k 's input is given by $L_{k,t} = (W_{k,t}/W_t)^{-\epsilon^w} L_t$.

I assume that a union faces a quadratic utility cost when adjusting its nominal wage $W_{k,t}$. This cost is given by $\frac{\chi^w}{2} (\pi_{k,t}^w)^2 L_t$, where $\pi_{k,t}^w = \dot{W}_{k,t}/W_{k,t}$ is the rate of nominal wage k inflation. Labor unions act according to an equal-weighted sum of household preferences. Unions are small and take as given households' policy functions. Their dynamic wage setting problem is thus given by

$$\max_{\pi_{k,t}^w} \mathbb{E}_0 \int_0^\infty e^{-\int_0^t (\rho_s + \zeta) ds} \left[\int u(c_t, h_t) g_t d(a, k, z) - \frac{\chi^w}{2} (\pi_{k,t}^w)^2 L_t \right] dt. \quad (1.9)$$

Importantly, union k directly controls the evolution of its wage path, so that there are no

stochastic innovations and nominal wages, $W_{k,t}$, are locally deterministic. Furthermore, unions are small and take as given the consumption policy function of the household, the demand function of the labor packer, as well as the distribution of households and its evolution. In the zero-inflation steady state, the real wage and marginal cost will be pinned down by $(1 - \tau^{\text{lab}})w \int u'(c)g = \frac{\epsilon^w}{\epsilon^w - 1}v'(H)$. See Appendix A.1 for additional details on the union problem.

1.3.5 Government policy and ZLB

In the baseline model, the scope of fiscal policy is limited. The government budget constraint is simply given by

$$r_t B^G + \tau_t^{\text{lump}} + \int \tau^{\text{UI}}(z)g_t(a, k, z)d(a, k, z) + G_t = \Pi_t + \tau^{\text{lab}}w_t L_t. \quad (1.10)$$

Fiscal expenditures include lump-sum rebates and unemployment insurance payments to households. The government is assumed to be a net debtor, with a constant level of riskfree debt outstanding given by B^G , on which it pays the real interest rate r_t . Finally, the government purchases the final consumption good at rate G_t . The government finances these expenditures with a labor income tax and with the corporate profits collected from the goods producing sector, Π_t .

Monetary policy in this model follows a simple Taylor rule and is subject to the zero lower bound (ZLB) constraint on nominal interest rates, so that

$$i_t = \max \left\{ r^* + \bar{\pi} + \lambda_\pi \pi_t + \lambda_Y y_t, 0 \right\}, \quad (1.11)$$

where r^* is the steady state riskfree rate, and $y_t = \log(Y_t/Y^*)$ denotes the output gap.

1.3.6 Aggregation and market clearing

I denote by $g_t(a, k, z)$ the cross-sectional household distribution at time t over liquid assets, a , illiquid assets, k , and employment status, z . Aggregation in this economy, for example of consumer expenditures, then takes the form

$$C_t = \int c_t(a, k, z)g_t(a, k, z)d(a, k, z).$$

The aggregate stocks of liquid and illiquid assets held by households are defined, respectively, as $A_t = \int a g_t(a, k, z) d(a, k, z)$ and $K_t = \int k g_t(a, k, z) d(a, k, z)$.

Market clearing. Formally, there are five markets in this economy that must clear at all times. The clearance of two of these markets, however, the rental market for capital and the labor market, is already implicit in the notation I have adopted: The aggregate stock of capital used in production must equal the aggregate stock of the illiquid asset held by households, and aggregate hours must equal the aggregate labor supply bundled by unions and the labor packer.

This leaves the bond market, which clears when the stock of liquid assets held by households equals the total riskfree debt issued by the government,

$$A_t = B^G.$$

In particular, as long as government debt is constant the stock of liquid assets is also constant. Second, the market for the final consumption good clears when

$$Y_t = C_t + I_t + \Phi_t + \Psi_t + G_t,$$

where C_t is aggregate consumption, $I_t + \Phi_t$ denotes gross investment expenditures including adjustment costs, and $\Psi_t = \int \psi(\iota_t, k) g_t(a, k, z) d(a, k, z)$ is the aggregate adjustment cost paid by households to trade in the illiquid asset.

This leaves, finally, the market for new capital, where households trade with the capital producing firm. All new capital produced must be purchased by households, and the price of capital Q_t adjusts to clear this market,

$$I_t = \int \iota_t(a, k, z) g_t(a, k, z) d(a, k, z).$$

In Appendix A.1.5, I provide an illustrative derivation of Walras' law.

Profits and the return on capital. What remains to be specified is the dissemination of profits and the composition of the rate of return on capital, dR_t . In the baseline model, I assume that pure profits earned by the goods producing sector are paid out uniformly to all households, via the government's lump-sum rebate τ_t^{lump} . Pure profits earned by the capital producing sector,

on the other hand, are disseminated in proportion to a household's illiquid asset holdings. That is, they constitute part of the return on capital.

Therefore, the real rate of return on capital is given by

$$dR_t = \left(r_t^k + \frac{\Pi_t^Q}{K_t} \right) dt \quad (1.12)$$

where r_t^k is the real rental rate paid by firms and Π_t^Q / K_t is the profit from capital production per unit of aggregate capital. The rate of return on the *stock* of capital features no capital gains term. If the household's portfolio equations are rewritten in terms of *net worth*, as is more common in the context of asset pricing, a capital gains term emerges that loads on the aggregate risk factor (see Appendix A.1.1 for details).

1.4 A Global Solution Method

Why are models like the one presented in Section 1.3 difficult to solve numerically? In any rational expectations equilibrium, agents must base their actions today on forecasts of future prices that are consistent with the economy's true law of motion. And in any heterogeneous-agent model, future prices will, through market clearing, depend on the future cross-sectional distribution of agents. Therefore, agents must consistently forecast the evolution of this distribution, which becomes part of the aggregate state space. The state space of the economy presented in Section 1.3 is therefore infinite-dimensional. Before moving on to my solution method, I will develop this argument formally and present the definition of a recursive equilibrium for my economy. The discussion in Section 1.4 is focused exclusively on the paper's methodological contribution. Readers who wish to proceed directly to the economic results may skip to Section 1.5 without loss.

The aggregate state space of my model is given by $\Gamma_t = (\rho_t, g_t)$, where ρ_t is the exogenous discount rate shock and g_t is the cross-sectional household distribution. The stationary value function of a household can be written as $V_t(a, k, z) = V(a, k, z, \Gamma_t)$, with slight abuse of notation. The household value function thus takes an infinite-dimensional, or measure-valued, input. While a, k and Γ are continuous state variables, z is discrete. I will therefore denote by

$V^j(a, k, \Gamma) = V(a, k, z_j, \Gamma)$ the value function of a household with employment status j and by $g^j(a, k, \rho) = g(a, k, z_j, \rho)$ the mass of households of employment type j with portfolio (a, k) in aggregate state Γ . Since the household distribution g is an argument of the value function, I will first characterize its evolution over time. For convenience, I introduce the shorthand notation $s^j(a, k, \Gamma)$ and $m^j(a, k, \Gamma)$ for the drift in liquid and illiquid portfolio positions, respectively.

The household distribution $g_t(a, k, z)$ evolves through time according to a Kolmogorov forward equation given by

$$\begin{aligned} \frac{dg^j(a, k, \rho)}{dt} = & -\partial_a \left[s^j(a, k, \Gamma) g^j(a, k, \rho) \right] - \partial_k \left[m^j(a, k, \Gamma) g^j(a, k, \rho) \right] \\ & - \lambda^j(\Gamma) g^j(a, k, \rho) + \lambda^{-j}(\Gamma) g^{-j}(a, k, \rho), \end{aligned} \quad (1.13)$$

where ∂_x denotes the partial derivative operator with respect to state variable x , and $-j$ denotes the employment type that *is not* j . See Appendix A.2.1. For compacter notation, equation (1.13) can also be written $(\mathcal{A}^* g^j)(a, k, \rho) \equiv \frac{d}{dt} g^j(a, k, \rho)$.²⁸

Formally, V^j solves a system of Hamilton-Jacobi-Bellman (HJB) equations given by

$$\begin{aligned} (\rho + \zeta) V^j(a, k, \Gamma) = & \max_{c^j, h^j, \nu^j} \left\{ u(c^j, h^j) + s^j \partial_a V^j(a, k, \Gamma) + m^j \partial_k V^j(a, k, \Gamma) \right\} \\ & + \lambda^j(\Gamma) \left[V^{-j}(a, k, \Gamma) - V^j(a, k, \Gamma) \right] + \theta_\rho (\bar{\rho} - \rho) \partial_\rho V^j(a, k, \Gamma) \\ & + \frac{\sigma_\rho^2}{2} \partial_{\rho\rho} V^j(a, k, \Gamma) + \underbrace{\sum_l \int \frac{\delta V^l}{\delta g} (\mathcal{A}^* g^l)(a, k, \rho) d(a, k)}_{\text{Effect of changes in cross-sectional distribution on household value function}} \end{aligned} \quad (1.14)$$

subject to the household budget, borrowing and short-sale constraints. Recall that s^j and m^j are only used as shorthand notation for the drift in the household's liquid and illiquid asset positions, respectively. Equation (1.14) is a characterization of the *stationary* household value function and therefore internalizes the effects of changes in the aggregate state of the economy. The last term in the third row captures the effect changes in the household cross-sectional distribution have on this value function. In particular, since g is infinite-dimensional, a functional Gateaux derivative

²⁸ \mathcal{A}^* is a functional operator that I will discuss further below.

δ is used.²⁹ This is precisely the term that cannot directly be implemented in any numerical scheme to solve this value function and must therefore be approximated.³⁰

We are now ready to define a recursive equilibrium of this economy. The definition of recursive equilibrium formally expresses the intuitive account with which I started this section: Since households must consistently forecast future aggregates, the infinite-dimensional cross-sectional distribution becomes part of their state space.³¹

Definition. (*Recursive Equilibrium*) A recursive competitive equilibrium of this economy is defined as the sets of functions $\{V^j, g^j, c^j, h^j, v^j\}(a, k, \Gamma)$ for $j \in \{E, U\}$ and $\{r, r^k, q, \tau^{\text{lump}}, H, w, Y, K, C, I, L, \pi, \pi^w\}(\Gamma)$ such that:

- (i) (Household optimization) $\{V^j, c^j, h^j, v^j\}$ solve the HJB (1.14) given all aggregates.
- (ii) (Firm and union optimization) Given aggregates, $\pi(\Gamma)$ solves the firm problem (1.5) and $\pi^w(\Gamma)$ solves the union problem (1.9) in each aggregate state Γ .
- (iii) (Aggregation) For each aggregate state Γ , the description of the aggregate household sector, i.e. $C(\Gamma), I(\Gamma)$ and $H(\Gamma)$, is consistent with aggregation from the micro level, using the

²⁹Formally, the value function $V(a, k, z, \rho, g)$ is defined on the product space comprised by the finite-dimensional state space for (a, k, z, ρ) and the space of measures for g . The partial differential equation (1.14) that characterizes V on this product space is also known as the “master equation” (see for example Cardaliaguet *et al.* (2015) and, originally, Lasry and Lions (2007) and Lions (2011)). The solution method I propose in this paper can therefore be viewed as a numerical implementation of the master equation. To my knowledge, this paper presents one of the first global numerical algorithms to solve the master equation of a mean field game with common noise.

³⁰To build intuition, imagine that the cross-sectional distribution was a vector $g = (g_1, \dots, g_N) \in \mathbb{R}^N$ rather than an infinite-dimensional function. This illustration is similar to one presented in Ahn *et al.* (2017). The HJB equation then becomes

$$\begin{aligned} (\rho + \zeta)V^j(a, k, g, \rho) &= \max_{c^j, h^j, v^j} \left\{ u(c^j, h^j) + s^j \partial_a V^j(a, k, g, \rho) + m^j \partial_k V^j(a, k, g, \rho) \right\} \\ &\quad + \lambda^j(g, \rho) \left[V^{-j}(a, k, g, \rho) - V^j(a, k, g, \rho) \right] + \theta_\rho (\bar{\rho} - \rho) \partial_\rho V^j(a, k, g, \rho) \\ &\quad + \frac{\sigma_\rho^2}{2} \partial_{\rho\rho} V^j(a, k, g, \rho) + \sum_{i=1}^N \sum_l \frac{\partial V^j(a, k, g, \rho)}{\partial g_i^l} g_i^l. \end{aligned}$$

The problematic term becomes a sum of partial derivatives, which can in principle be implemented numerically since everything is now finite-dimensional. One contribution of my solution method is to generalize this approach to any arbitrary finite-dimensional approximation of g .

³¹One could also define a sequence equilibrium for this economy. For example, Brunnermeier and Sannikov (2014) define a sequence equilibrium for an economy that is structurally similar to a variant of mine in which the distribution is a degenerate two-point delta function.

cross-sectional distribution $g^j(a, k, \Gamma)$.

- (iv) (General equilibrium) All markets clear, and the remaining macroeconomic aggregates solve the model's general equilibrium conditions.
- (v) (Distribution and rational expectations) The cross-sectional household distribution evolves according to (1.13) and household behavior is based on forecasts that are consistent with the true law of motion of the aggregate state of the economy.

1.4.1 Finite-dimensional distribution representations

Any numerical solution of a heterogeneous-agent model with aggregate risk must approximate the aggregate state space in a finite-dimensional subspace. In other words, any numerical solution method will implement an approximate value function that takes as an input a finite-dimensional approximation of the distribution g_t . In this paper, I focus on the class of finite-dimensional representations given by

$$F(\alpha_t)(\mathbf{x}) \approx g_t(\mathbf{x}), \tag{1.15}$$

where $\hat{g}_t(\mathbf{x}) = F(\alpha_t)(\mathbf{x})$ is then interpreted as the approximate cross-sectional distribution parameterized by α_t . For compact notation, I use $\mathbf{x}_t = (a_t, k_t, z_t)$ to denote the vector of idiosyncratic state variables. While the representation (1.15) is quite general, it is most illustrative to think of F as a set of basis functions defined over \mathbf{x} that are parameterized by α_t . Before proceeding, I present an example for illustration.³²

Example 2. Denote by $T^n(x)$ the n th Chebyshev polynomial over x . We can stack these basis functions in a row-vector denoted $T(\mathbf{x})$. Letting $\alpha_t \in \mathbb{R}^N$ denote the column-vector of coefficients, equation (1.15) becomes

$$F(\alpha_t)(\mathbf{x}) = T(\mathbf{x})\alpha_t = \sum_{n=1}^N \alpha_t^n T^n(\mathbf{x}) \approx g_t(\mathbf{x}).$$

³²In the context of a local solution, Winberry (2020) uses an approximate distribution representation that also takes the form (1.15). In this sense, my paper can be viewed as taking an approach similar to Winberry (2020) but in the context of a global solution method. I present several analytical and numerical tools below that help overcome the curse of dimensionality that oftentimes becomes a hurdle for global solution methods.

While the set of selected basis functions F remains constant over time, time variation in the parameters α_t allows to capture approximately the evolution of the true cross-sectional distribution g_t over time. In the context of the above example: The researcher picks the N Chebyshev polynomials $T^n(x)$ ex ante and they don't change over the course of a simulation. The basis function coefficients α_t^n , on the other hand, are time-varying. In particular, we will want to specify the time variation in α_t^n so as to match the time variation in $g_t(x)$ as closely as possible.

Approximate economy. How does the approximation (1.15) make the model of Section 1.3 tractable? I will refer to the model solved under the finite-dimensional representation (1.15) as an *approximate economy*. That is, an approximate economy uses $\hat{g}_t(x) = F(\alpha_t)(x)$ as its cross-sectional distribution.

Recall the source of intractability in the household's value function equation (1.14): As part of the aggregate state space, the infinite-dimensional distribution g_t enters V as an argument. The aggregate state space of the approximate economy no longer features the true cross-sectional distribution but rather its finite-dimensional approximation, $\hat{g}_t(x)$. That is, the *approximate aggregate state space* becomes $\hat{\Gamma}_t = (X_t, \hat{g}_t)$. Under the representation $\hat{g}_t(x) = F(\alpha_t)(x)$, we can further simplify the aggregate state space: All information we require to track the evolution of $\hat{g}_t(x)$ over time is equivalently encoded in α_t . Therefore, the approximate aggregate state space becomes $\hat{\Gamma}_t = (X_t, \alpha_t)$.³³

Reducing the aggregate state space from Γ_t to $\hat{\Gamma}_t$ makes the household problem (1.14) tractable. The value function that characterizes household behavior in the approximate economy is now given by

$$\begin{aligned} V(x, \rho, g) &\approx V(x, \rho, \hat{g}) \\ &= \hat{V}(x, \rho, \alpha). \end{aligned}$$

³³My method also builds on previous work that makes use of *distribution selection functions* (e.g. Algan *et al.* (2008) and Reiter (2010)). Like the original Krusell-Smith algorithm, these papers use moments to characterize the distribution. In the spirit of my approach, however, they postulate a mapping like (1.15) that is consistent with their moments. This distribution selection function associates a cross-sectional distribution with each possible realization of moments. These algorithms share several important features with the method I propose here, and I show in the Numerical Appendix that several of my results below can also be useful for those algorithms. For example, I show how to modify the Algan *et al.* (2008) algorithm so that it can directly make use of Proposition 3.

In particular, the household value function now only takes finite-dimensional arguments. In the approximate economy households only have to forecast the approximate distribution $\hat{g}_t(\mathbf{x})$ or, equivalently, the new state variables α_t . I show below that, under quite general conditions, equilibrium in the approximate economy is characterized by the law of motion

$$d\alpha = \mu_\alpha(\hat{\Gamma})dt + \sigma_\alpha(\hat{\Gamma})dB.$$

That is, α_t follows an Ito diffusion process, and μ_α and σ_α are equilibrium objects that we must solve for. In an approximate economy based on (1.15), therefore, household behavior is now characterized by

$$\begin{aligned} (\rho + \zeta)\hat{V}^j(a, k, \hat{\Gamma}) = & \max_{\hat{c}^j, \hat{h}^j, \hat{p}^j} \left\{ u(\hat{c}^j, \hat{h}^j) + \hat{s}^j \partial_a \hat{V}^j(a, k, \hat{\Gamma}) + \hat{m}^j \partial_k \hat{V}^j(a, k, \hat{\Gamma}) \right\} \\ & + \hat{\lambda}^j(\hat{\Gamma}) \left[\hat{V}^{-j}(a, k, \hat{\Gamma}) - \hat{V}^j(a, k, \hat{\Gamma}) \right] + \theta_\rho(\bar{\rho} - \rho) \partial_\rho \hat{V}^j(a, k, \hat{\Gamma}) \\ & + \frac{\sigma_\rho^2}{2} \partial_{\rho\rho} \hat{V}^j(a, k, \hat{\Gamma}) + \underbrace{\mu_\alpha(\hat{\Gamma}) \partial_\alpha \hat{V}^j(a, k, \hat{\Gamma}) + \frac{1}{2} \sigma_\alpha(\hat{\Gamma})^T \partial_{\alpha\alpha} \hat{V}^j(a, k, \hat{\Gamma}) \sigma_\alpha(\hat{\Gamma})}_{\text{Effect of cross-sectional distribution on household value function in approximate economy}}. \end{aligned} \quad (1.16)$$

The new terms in the third row are now readily computable given the functions $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$.

Working with a finite-dimensional distribution representation of the form (1.15) offers numerous advantages.³⁴ While this approach is commonly used in the context of local perturbation methods, it has proven intractable in the context of global methods due to the curse of dimensionality. That is, accurate approximations $\hat{g}_t(\mathbf{x}) = F(\alpha_t)(\mathbf{x})$ oftentimes require high-dimensional α_t . In the remainder of this section, I discuss several analytical and numerical tools that help overcome this challenge. In the interest of brevity and accessibility, I only sketch technical arguments in the main text. Details are provided in Appendix A.2 and a separate

³⁴A key advantage of being able to work with an approximate distribution object $\hat{g}_t(\mathbf{x})$ on the grid is that it allows us to evaluate market clearing conditions. If we instead approximate the cross-sectional distribution with moments, this is generally not possible. For example, to evaluate the goods market clearing condition in Section 1.3 requires first evaluating aggregate household consumption, $\hat{C}(\Gamma) = \int \hat{c}(\mathbf{x}, \Gamma) \hat{g}(\mathbf{x}, \Gamma) d(\mathbf{x})$, which is only possible given \hat{g} . With a distribution object “on the grid”, we are able to solve for prices and other general equilibrium objects directly on the grid. By contrast, algorithms like Krusell and Smith (1998) must resort to a costly “simulate-estimate” approach, which I describe below, to solve for these general equilibrium objects.

1.4.2 How to overcome the curse of dimensionality?

For most economic applications of interest, accurate approximations of the distribution will require high-dimensional $F(\alpha_t)(x)$. Global methods will therefore quickly encounter the curse of dimensionality. In Schaab and Zhang (2020), we have developed an adaptive sparse grid library for solving partial differential equations in continuous time. Sparse grid methods aim to combat the curse of dimensionality when representing functions on high-dimensional grids. Regular sparse grids reduce the complexity of a grid in d dimensions from $\mathcal{O}(n^d)$, where n denotes the number of grid points per dimension, to $\mathcal{O}(n \log(n)^{d-1})$ (see for example Bungartz and Griebel (2004)). They have found occasional use in economics (see for example Krueger and Kubler (2004) and Judd *et al.* (2014)). Truly adaptive sparse grids were first introduced to economics by Brumm and Scheidegger (2017) but had previously enjoyed more popularity in physics and applied math.

Grid adaptation leverages the insight that not all grid points are “equally valuable” when representing a function on a grid. A rule of thumb is that an accurate function representation requires more grid points in areas where the function is particularly concave. A linear function in one dimension, for example, can be represented perfectly using only two grid points and linear interpolation elsewhere. Adaptive sparse grid algorithms automatically add and drop grid points to maximize the efficiency of grid point placement for a given application.

The main contribution of Schaab and Zhang (2020) relative to Brumm and Scheidegger (2017) is to develop a robust adaptive sparse grid infrastructure for solving (partial) differential equations in continuous time. While a self-contained description of our library is beyond the scope of this paper, I will illustrate the power of adaptive sparse grids in the context of the model of Section 1.3. For simplicity, I focus on the model’s deterministic steady state, formally defined as $V^{j,0}(\cdot) = \lim_{\sigma_\rho \rightarrow 0} V^j(\cdot; \sigma_\rho)$. Figure 1.1 displays fully converged solutions of

³⁵An important question is under what conditions the scheme in (1.15) is a *consistent* approximation to the system of coupled PDEs, given by the HJB and KF equations, that characterizes the true model. In other words, under what conditions does a sequence of approximating economies characterized by $\{F^n(\alpha_t^n)\}_n$ converge to the true economy as $n \rightarrow \infty$. I am currently working on proving such a convergence result for particular classes of basis functions F .

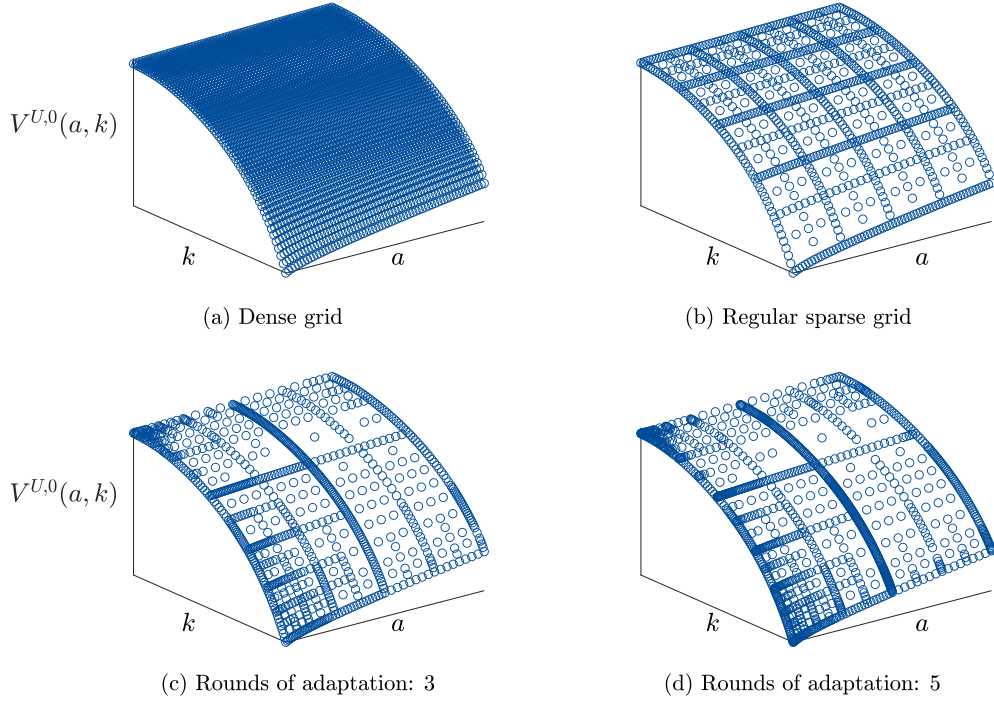


Figure 1.1: Household value function on increasingly adapted sparse grids

Notes. Each panel displays the value function of an unemployed household in the deterministic steady state, i.e. $V^{U,0}(a, k) = \lim_{\sigma_p \rightarrow 0} V^U(a, k; \sigma_p)$. Panel (a) solves V on a dense grid consisting of a full set of equidistantly spaced grid points. Panel (b) solves V on a regular sparse grid which removes grid points from the dense grid according to a prespecified pattern. Panels (c) and (d) display the converged value function on increasingly adapted sparse grids.

$V^{U,j}(a, k)$, an unemployed household's value function in the deterministic steady state, across four increasingly adapted grids.³⁶ A dense grid consists of a full set of equidistantly spaced grid points. The key insight of sparse grid methods is that not all of these grid points are equally valuable in the representation of V . In Panel (b), the value function is solved on a so-called regular sparse grid, which starts from a dense grid and removes grid points according to a prespecified pattern. Panels (c) and (d), finally, adapt the grid and place grid points efficiently to conform to the concavity in V . Panel (d) illustrates clearly that grid points are placed in the region where a and k are low and household behavior at the micro level exhibits non-linearities. While Figure 1.1 demonstrates that sparse grid methods are already useful for two-dimensional

³⁶For illustration, Figure 1.1 uses more grid points in the idiosyncratic household dimensions than I use to solve the quantitative model in higher dimensions when $\sigma_p > 0$.

grids, they become especially powerful in higher dimensions.

1.4.3 How to find a consistent law of motion for α_t ?

To solve the household problem in the approximate economy, we must solve for the equilibrium law of motion of α_t . This is highlighted by equation (1.16) which depends directly on μ_α and σ_α .

The traditional approach to solving for the law of motion of α_t is as follows: (i) Start with a guess for $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$. (ii) Given this guess, solve equation (1.16) and the rest of the model on the aggregate state space. (iii) Simulate the model. (iv) Update the original guess for $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$ using estimates from the simulated data, and start again at step (i). These steps are repeated until convergence. This strategy was popularized by Krusell and Smith (1998) and remains an integral part of most global solution methods currently in use.

While this “simulate-estimate” approach is, of course, valid and oftentimes useful in algorithms based on (1.15), it is also very costly.³⁷ In practice, it typically requires hundreds of (outer) iterations in the fixed point algorithm sketched above to consistently estimate the law of motion of α_t . In each step, the entire model must be solved, including the high-dimensional household value function, and simulated.

An advantage of an algorithm based on (1.15) is that it offers a more efficient alternative to find this law of motion. Indeed, Proposition 3 below provides analytical formulas for $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$ that can be easily computed. In this sense, and in sharp contrast to the Krusell and Smith (1998) algorithm, finding the consistent law of motion incurs almost no increase in numerical complexity.

To develop intuition for this result, I present a heuristic but hopefully illustrative derivation for the special case where $\sigma_\alpha(\hat{\Gamma}) = 0$. This turns out to be true for the model of Section 1.3. Proposition 3 below provides more general formulas for $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$, which are developed formally in Appendices A.2.1 through A.2.3.

According to equation (1.13), the true law of motion of the distribution can be written as

³⁷In practice, I use a combination of the “simulate-estimate” approach and Proposition 3 to solve the quantitative model of Section 1.3. In the Numerical Appendix, I show that exclusively using Proposition 3 can lead to significant performance gains in simpler models such as the Krusell and Smith (1998) model or a one-asset HANK model.

$dg_t(\mathbf{x}) = -\partial_a(sg) - \partial_k(mg) - \partial_z(\mu_z g) \equiv (\mathcal{A}^*g_t)(\mathbf{x})dt$. Under the approximation $\hat{g}_t(\mathbf{x}) \approx g_t(\mathbf{x})$, we have $d\hat{g}_t \approx dg_t$. The law of motion of the approximate cross-sectional distribution can therefore also be written as

$$d\hat{g}_t(\mathbf{x}) = (\mathcal{A}^*\hat{g}_t)(\mathbf{x})dt \quad (1.17)$$

where, abusing notation slightly, \mathcal{A}^* is now evaluated in the approximate economy. Conjecturing that $d\alpha_t = \mu_\alpha(\hat{\Gamma})dt$ and differentiating $\hat{g}_t(\mathbf{x}) = F(\alpha_t)(\mathbf{x})$ with respect to time, we have

$$d\hat{g}_t(\mathbf{x}) = F_\alpha(\alpha_t)(\mathbf{x})\mu_\alpha(\hat{\Gamma})dt$$

where F_α is the gradient of F with respect to α . Using equation (1.17) and matching coefficients, we arrive at a functional equation that characterizes $\mu_\alpha(\hat{\Gamma})$ via

$$F_\alpha(\alpha_t)(\mathbf{x})\mu_\alpha(\hat{\Gamma}) = (\mathcal{A}^*F(\alpha_t))(\mathbf{x})$$

To “invert” this equation and solve for $\mu_\alpha(\hat{\Gamma})$ we must assume a specific (estimation) norm that we want to minimize. Under the $\mathbb{L}^2(\mathbf{x})$ norm, we arrive at $\mu_\alpha = (F_\alpha^T F_\alpha)^{-1} F_\alpha^T (\mathcal{A}^*F)$.

Proposition 3 is presented for the choice of norm $\mathbb{L}^2(\mathbf{x})$, and under the additional assumption that x_t follows a diffusion process.

Proposition 3. *Let $\hat{\Gamma}_t = \{\rho_t, \alpha_t\}$ denote the aggregate state of an approximate economy under a distribution representation (1.15). Then:*

1. *The law of motion of α_t is given by*

$$d\alpha = \mu_\alpha(\hat{\Gamma})dt + \sigma_\alpha(\hat{\Gamma})dB \quad (1.18)$$

2. *The choice of μ_α and σ_α that minimizes forecast errors in the \mathbb{L}^2 -norm is given by*

$$\mu_\alpha = (F_\alpha^T F_\alpha)^{-1} F_\alpha^T \left[\mathcal{A}^*F - \frac{1}{2} \sigma_\alpha^T F_{\alpha\alpha} \sigma_\alpha \right] \quad (1.19)$$

and

$$\sigma_\alpha = (F_\alpha^T F_\alpha)^{-1} F_\alpha^T \mathcal{B}^* F, \quad (1.20)$$

where F_α and $F_{\alpha\alpha}$ are the Jacobian and Hessian of F , respectively, \mathcal{A}^* is the adjoint of the infinitesimal operator defined by the HJB equation, and \mathcal{B}^* is the diffusion coefficient in the stochastic

Kolmogorov forward equation for g_t .

Proposition 3 promises a shortcut in the design of algorithms to solve heterogeneous-agent models with aggregate risk: Equations (1.19) and (1.20) provide formulas to compute the internally consistent law of motion of α_t directly.³⁸ The most surprising and useful aspect of Proposition 3 is that each term that features in equations (1.19) and (1.20) is readily computable inside the household's value function iteration step. For illustration, consider applying Proposition 3 to the model of Section 1.3, which yields $\sigma_\alpha(\hat{\Gamma}) = 0$ (as conjectured above) and

$$\mu_\alpha = -(F_\alpha^T F_\alpha)^{-1} F_\alpha^T \left[\partial_a(\hat{s}F) + \partial_k(\hat{m}F) + \partial_z(\mu_z F) \right]. \quad (1.21)$$

For a given choice of $F(\cdot)$ and with α on the grid as an aggregate state variable, the terms $F(\alpha)$, F_α and $F_{\alpha\alpha}$ can be computed using standard numerical derivatives even before the value function iteration step. In many relevant cases, F_α and $F_{\alpha\alpha}$ will, in fact, be available as closed form analytical expressions. All remaining terms in equation (1.21) are comprised either of household policy functions, \hat{s} and \hat{m} , or the employment transition rates, μ_z or equivalently λ^j .³⁹

Algorithm structure. In light of these observations, I propose to compute $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$ directly as part of the household value function iteration (VFI) step. My algorithm therefore adds an additional step to an otherwise standard VFI procedure. The k th iteration follows these four steps:

1. Start with this iteration's guess \hat{V}^k .
2. Compute policy functions \hat{s}^k and \hat{m}^k using \hat{V}^k and the household's FOCs.
3. Use \hat{s}^k and \hat{m}^k to compute $\mu_\alpha^k(\hat{\Gamma})$ and $\sigma_\alpha^k(\hat{\Gamma})$ according to equations (1.19) and (1.20).
4. Use (1.16) to find \hat{V}^{k+1} , and repeat from step (1.) until convergence.

³⁸Ahn *et al.* (2017) make a similar argument. Relative to their paper, Proposition 3 applies to settings where the cross-sectional distribution of agents itself is stochastic, in which case $\mathcal{B}^* \neq 0$ and $\sigma_\alpha(\hat{\Gamma}) \neq 0$. Formally, I nest models in which the distribution's law of motion is given by a stochastic rather than an ordinary partial differential equation. Asset pricing models with portfolio choice problems are typically of this kind.

³⁹In Appendix A.2.2, I show that \mathcal{A}^* and \mathcal{B}^* are, more generally, comprised of household policy functions and transition rates.

Notice that, in iteration k of this modified VFI algorithm, $\mu_\alpha^k(\hat{\Gamma})$ and $\sigma_\alpha^k(\hat{\Gamma})$ are only required to evaluate (1.16) in step (4.) and update the value function \hat{V}^{k+1} . In other words, equations (1.19) and (1.20) are computed *after* the household policy functions and can therefore be readily evaluated. Taking general equilibrium prices as given, this modified VFI algorithm solves for both \hat{V} and an internally consistent law of motion for α_t at the same time.⁴⁰ By contrast, the traditional approach that follows Krusell and Smith (1998) constructs an outer fixed point to solve for μ_α and σ_α .

1.4.4 How to Choose Efficient Representations $F(\cdot)$?

The next natural question is how F should be chosen in practice. One straightforward approach is to choose F directly from a parametric family.⁴¹ When the functional form $F(\alpha)$ is chosen ex ante in this way, no simulation step is required at any point in the algorithm, which is one of its most compelling features.⁴²

The approximation mapping (1.15) can also be implemented non-parametrically. The non-parametric approach can deliver large performance gains, especially when \mathbf{x}_t is high-dimensional, but comes at the cost of an outer simulation step. For illustration, I will discuss the special case where $F(\cdot)$ is affine in $\alpha_t \in \mathbb{R}^N$, that is $F(\alpha_t)(\mathbf{x}) = C(\mathbf{x}) + T(\mathbf{x})\alpha_t$ for some functions C and T . In practice, I simply set $C(\mathbf{x})$ equal to the distribution associated with the deterministic steady state, that is $C(\mathbf{x}) = g^0(\mathbf{x}) \equiv \lim_{\sigma_\rho \rightarrow 0} g(\mathbf{x}; \sigma_\rho)$. As in Example 2, $T(\cdot)$ should be interpreted as a row-vector of basis functions and α_t as the column-vector of coefficients.

⁴⁰What is particularly nice about applying Proposition 3 in practice is that it prescribes a method for computing $\mu_\alpha(\hat{\Gamma})$ and $\sigma_\alpha(\hat{\Gamma})$ that is independent from the researcher's choice of F . In practice, therefore, the researcher can switch out different candidates for F with ease to determine which representation works best. In other currently popular global solution methods, the algorithm step to find the law of motion of α_t (or other moments) is oftentimes highly dependent on the particulars of the distribution approximation. It has therefore been difficult in practice to switch between and compare different solution methods (e.g. candidates for F) seamlessly.

⁴¹See Appendix A.2.3 for details. Parametric families that are commonly used include hat functions (nodal basis), Chebyshev polynomials, radial basis functions, generalized beta density functions and splines.

⁴²The complexity of solving a model like the seminal Krusell and Smith (1998) benchmark is therefore reduced to a single value function iteration step that jointly computes \hat{V} and the law of motion of α_t (see previous subsection). When the model requires an additional general equilibrium fixed point to solve for prices, the solution complexity is reduced to a single iteration of this fixed point.

Algorithm structure. The non-parametric approach constructs an outer fixed point whose n th iteration follows three steps:

1. Given this iteration's distribution representation,

$$F^n(\alpha_t)(\mathbf{x}) = g^0(\mathbf{x}) + \sum_{k=1}^n T^k(\mathbf{x})\alpha_t^k, \quad \alpha_t \in \mathbb{R}^n$$

solve the n th *approximate economy* that uses $\hat{g}_t^n(\mathbf{x}) = F^n(\alpha_t)(\mathbf{x})$ as distribution.⁴³

2. Given a solution of approximate economy n , simulate the model to obtain simulated data for the distribution, $g_t^{n,\text{sim}}$. Compute the forecast error $F^n(\alpha_t)(\mathbf{x}) - g_t^{n,\text{sim}}(\mathbf{x})$.
3. Find the basis function $T^{n+1}(\mathbf{x})$ that minimizes forecast errors under a prespecified norm, subject to equation (1.18) for α_t . Update F^{n+1} and iterate until a desired accuracy is reached.

Intuitively, step (3.) looks for a kind of functional principal component with explanatory power for the forecast residual. At the cost of having to simulate the model, this non-parametric approach delivers substantial efficiency gains especially when the idiosyncratic state space of agents is high-dimensional.⁴⁴ To achieve a given level of accuracy, the number of basis functions n required in this non-parametric approach is often far smaller than in the parametric approach. In Appendix A.2.4, I provide additional details, including on the choice of forecast error norms and the resulting estimation optimality conditions.

1.5 Taking the Model to the Data

In this section, I calibrate the model and confront its main predictions with data. The model is calibrated on a quarterly frequency, based on U.S. data.

Table 1.1: List of Calibrated Parameters

Parameters	Value	Target / Source
<i>Preferences</i>		
$\bar{\rho}$	Discount rate (p.q.)	2.2 % Riskfree rate
γ	Relative risk aversion	2 Standard
η	Inverse Frisch elasticity	2 Standard
ζ	Deathrate	0.556 % Average life span 45 years
<i>Household portfolio choice</i>		
\underline{a}	Borrowing constraint	-1 1.244 q average income
δ	Capital depreciation (p.q.)	1.25 %
ψ_0	Linear adjustment cost	0.044 Top 10% (liquid) wealth share
ψ_1	Convex adjustment cost	0.956 % households with $a \leq 0$ and $k > 0$
<i>Firms</i>		
α	Capital share	0.38
κ	Aggregate capital adjustment cost	40 $\Delta I / \Delta Y$ after $\Delta \sigma_\rho$
<i>Nominal rigidities</i>		
$\frac{\epsilon^f}{\epsilon^f - 1}$	Elasticity of substitution	1.10 CEE (2005)
χ^f	Price adjustment cost	350 ACEL (2011)
$\frac{\epsilon^w}{\epsilon^w - 1}$	Elasticity of substitution	1.05 CEE (2005)
χ^w	Avg. duration of wage contracts	0 Flexible-wage limit
<i>Government</i>		
B^G	Government debt outstanding	0.06 Average aggregate MPC
G	Government spending	0
λ_π	Taylor rule weight on inflation	1.5 Standard
λ_Y	Taylor rule weight on output	0.5 Standard
τ^{lab}	Income tax rate	0.2
τ^{UI}	Unemployment insurance	0.20 Chodorow-Reich and Karabarbounis (2016)
<i>Aggregate uncertainty</i>		
σ_ρ	Volatility: discount rate shock	0.003 Volatility of GDP growth
θ_ρ	Persistence: discount rate shock	0.22 Autocorrelation of GDP growth

1.5.1 Calibration strategy

Table 1.1 provides a summary of all model parameters except for the employment transition rates that are discussed in Section 1.5.2. Households have CRRA preferences in consumption and leisure, with a relative risk aversion coefficient of $\gamma = 2$, and an inverse Frisch labor supply

⁴³This step requires solving the household problem (1.16), the union problem and the firm problem, as well as finding the general equilibrium prices that clear all markets. As discussed in the previous subsection, finding an internally consistent law of motion for α_t is subsumed in the modified VFI step.

⁴⁴In the context of algorithms that work with a set of moments to approximate the distribution $g_t(x)$, the fixed point I describe is akin to the idea of successively adding moments to improve forecast accuracy. In my setting, this approach is complicated by the flexibility of being able to choose any $F(\cdot)$.

elasticity of $\eta = 2$. I calibrate $\bar{\rho}$, households' quarterly discount rate in the risky steady state, to match a natural rate of interest of roughly 2% (2.14% in the model) in the risky steady state. The deathrate ζ is calibrated to imply an average life span of 45 years, corresponding roughly to the average working life.

Households' portfolio choice is primarily affected by the adjustment cost on illiquid investments and the borrowing constraint on liquid assets. I set the latter to $\underline{a} = -1$, implying that households have access to roughly one quarter of average income in unsecured borrowing as in Kaplan *et al.* (2018). I calibrate the two parameters of the adjustment cost function to match the top 10% wealth share in liquid wealth (86% in both data and model), as well as the share of households with non-positive liquid net worth but a positive illiquid asset position (30% in the data, 27% in the model's risky steady state). In calibrating these parameters, I stay as close as possible to Kaplan *et al.* (2018) since the wealth distribution is not the primary focus of my contribution.

I set depreciation to 1.25% per quarter and the capital income share α to 0.38. I calibrate the aggregate capital adjustment cost parameter, κ , to match the business cycle sensitivity of investment, relative to output. In particular, the peak decline of aggregate investment following a fundamental risk shock is currently three times as large as that of output while aggregate consumption is less responsive.

The parameters governing nominal price rigidity are set equal to literature benchmarks. In particular, I set the elasticity of substitution so that $\frac{\epsilon^f}{\epsilon^f - 1} = 1.1$ as in Christiano *et al.* (2005), and I set χ^f to imply an average price duration of roughly 7 quarters. Altig *et al.* (2011) estimate an average price duration of 9.36 quarters. I solve the baseline model in the limit as wages are fully flexible, $\chi^w \rightarrow 0$.⁴⁵

The baseline model does not feature government spending, and the stock of government debt outstanding is constant. One weakness of the model with only two earnings types is that it is difficult to jointly match high MPCs and a high stock of liquid assets. In particular, matching an aggregate household MPC of roughly 15% requires setting the stock of outstanding

⁴⁵The cost of assuming flexible wages is, of course, the standard problem that the profits of goods producing firms are counter-cyclical. In the next iteration of the paper, I plan to allow for both price and wage rigidity.

government debt close to zero ($B^G = 0.06$). I set the income tax rate to 20%, and the effective unemployment insurance replacement rate to 20%. This is consistent with recent estimates accounting for take-up in Chodorow-Reich and Karabarbounis (2016). The central bank follows a Taylor rule with the standard literature weights on inflation and output of, respectively, 1.5 and 0.5.

Finally, I calibrate the volatility and persistence of the discount rate shock process to match the business cycle moments of GDP growth, with a quarterly auto-correlation of 0.85 and a standard deviation of 0.017.

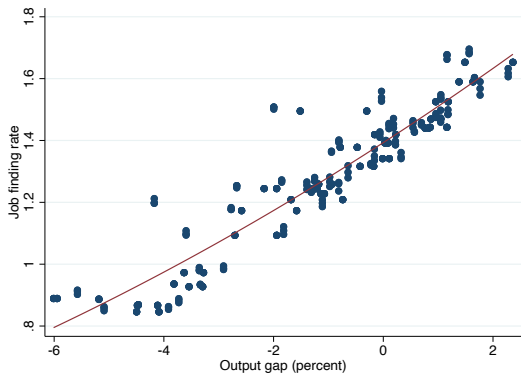
1.5.2 The ins and outs of unemployment

The employment transition rates faced by households play a central role in the quantitative results of this paper. I use microdata on employment transitions from the Current Population Survey (CPS) to estimate the sensitivity of the job finding and separation rates over the business cycle.⁴⁶ I use the resulting reduced-form estimates for $\lambda^j(\cdot)$ directly in the model. For additional details, see Appendix A.3.1.

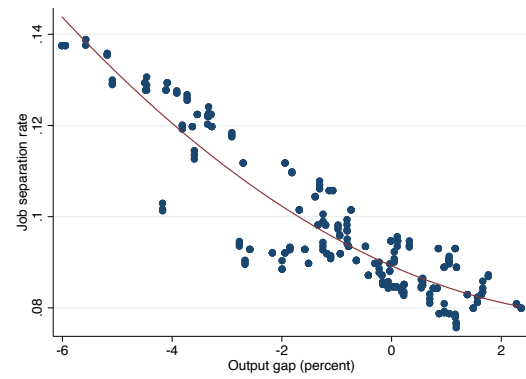
Exploiting the rotating panel nature of the CPS, I match household observations across subsequent months to construct monthly gross flows into and out of employment in a sample from 1996 to 2019. I employ a broad notion of non-employment including unemployed and marginally attached households, as well as those in involuntary part-time employment for economic reasons. This definition corresponds to the U6 unemployment rate. I show that these groups also exhibit sizable transition rates into and out of employment, which, following the logic of my model, motivates their inclusion.

Having constructed matched gross flow data at a monthly frequency, I use a continuous-time conversion formula to compute exact data counterparts to the Poisson arrival rates λ_t^E and λ_t^U in the model. In the spirit of Shimer (2012), this conversion also addresses a time aggregation bias in discrete-time flow data. Panels (c) and (d) of Figure 1.2 display the data time series for

⁴⁶There is a long tradition of constructing gross employment flows from matched CPS micro data. See for example Marston *et al.* (1976), Abowd and Zellner (1985), Darby *et al.* (1985), Darby *et al.* (1986), Poterba and Summers (1986), Blanchard *et al.* (1990), Shimer (2005), Fujita and Ramey (2009), Elsby *et al.* (2009), and Shimer (2012). In particular, I closely follow Shimer (2012) in the construction of monthly gross flow data.



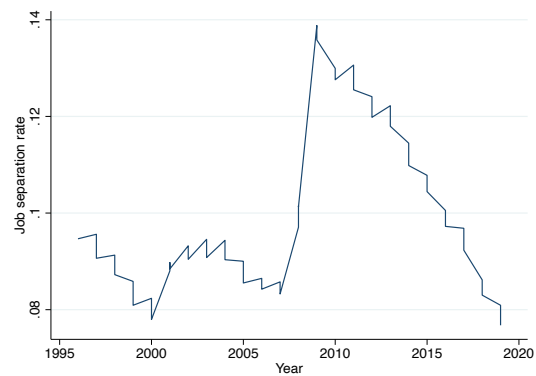
(a)



(b)



(c)



(d)

Figure 1.2: *Employment Transition Rates over the Business Cycle*

Notes. The raw CPS micro data for job finding and separation rates from 1996 to 2019 are displayed in Panels (c) and (d). The data are expressed as instantaneous, quarterly transition rates after a conversion from monthly gross flow data to continuous time. The underlying definition of unemployment used is broader than U3 headline unemployment and corresponds closely to U6. Panels (a) and (b) scatter the raw data against output gap estimates. The solid red lines correspond to a quadratic fit.

these job finding and separation rates.

Since these data objects correspond exactly to the instantaneous transition rates (at quarterly frequency) faced by households in the model, I use an external calibration approach to specify how λ_t^E and λ_t^U move over the business cycle. In the baseline calibration, I use the shortcut specification that the job finding and separation rates are only functions of the output gap, so

that $\lambda_t^j = \lambda^j(y_t)$. I implement this specification in the data using a Taylor expansion

$$\text{Job finding rate : } \lambda_t^U = 1.39 + 0.115 y_t + 0.0026 y_t^2 + \dots \quad (1.22)$$

(114.96) (13.66) (1.08)

$$\text{Job separation rate : } \lambda_t^E = 0.89 - 0.0053 y_t + 0.0006 y_t^2 + \dots \quad (1.23)$$

(88.67) (-6.85) (3.62)

and report the associated T-statistics after a Newey-West correction. Panels (a) and (b) of Figure 1.2 display the associated scatter plots. There is strong evidence of systematic time variation in employment transition rates over the business cycle. The job finding (separation) rate falls (increases) in bad times. Similarly, equation (1.23) shows that at least the job separation rate exhibits a non-linearity that is statistically significant.

In my baseline calibration, I use equations (1.22) and (1.23) directly to specify λ^j as a function of the economy's aggregate state.⁴⁷

1.5.3 Business cycle fluctuations

The model matches the asymmetry of business cycles in U.S. postwar history: recessions have been deeper and more pronounced in the data than economic booms.

My calibration only targets the volatility of GDP growth (0.017 in the data and 0.014 in the model) and the auto-correlation of GDP growth (0.85 in the data and 0.88 in the model). The model also matches other untargeted business cycle moments at least qualitatively. In Section 1.6.1, I show that demand shocks generate co-movement in output, consumption, investment and hours in the model. In particular, aggregate investment is more volatility and responsive than output, while aggregate consumption is less responsive.

Qualitatively, the model also matches higher moments of key business cycle aggregates. Figure 1.3 compares the distribution of GDP growth realizations in model simulations (yellow bars) against U.S. postwar data (blue bars). The distribution of GDP growth in U.S. data

⁴⁷The empirical transition rate series I have constructed do not exactly aggregate back up to the aggregate unemployment rate (U6) observed in the data. To ensure that the model matches the average unemployment rate, I adjust the constant term in equation (1.22). That is, my model matches the average job separation rate and the average U6 unemployment rate, as well as the cyclicalities exhibited by the job finding and separation rates in the sample from 1996 to 2019.

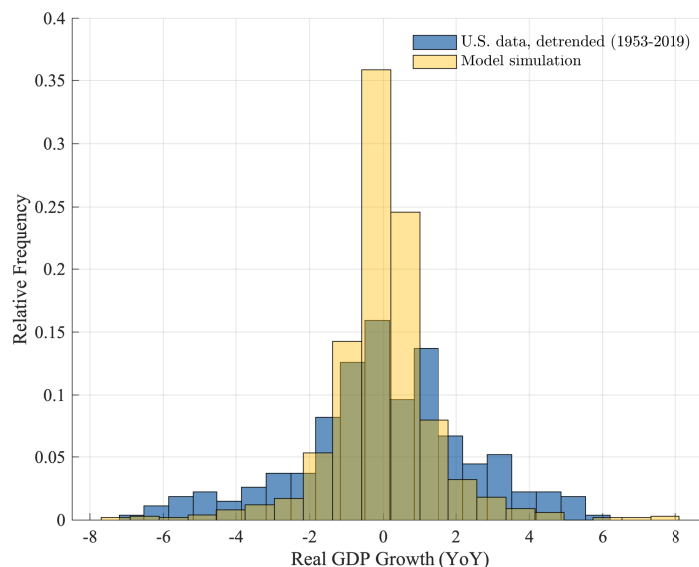


Figure 1.3: *Distribution of GDP growth in U.S. postwar data and model simulations*

Notes. Relative frequency histogram of distribution of year-over-year GDP growth in U.S. postwar data since 1953 (blue bars) and model simulations (yellow bars). Data is detrended using a Kalman filter.

since 1953 has fatter tails than in my model simulations. Nonetheless, output, consumption, investment and hours in model simulations exhibit significant negative skewness and positive kurtosis. While the skewness and kurtosis generated by the model are larger than their respective data counterparts, this is not unexpected in a meaningful sense. While the ZLB has only become a binding constraint for U.S. monetary policy quite recently, most recessions in my model simulations drive the economy to the ZLB. The model is therefore intended to represent this more recent episode in U.S. economic history, characterized by a structurally much lower natural rate of interest. Indeed, if I recompute the U.S. business cycle moments over a sample of the Great Moderation, starting in 1980, the implied skewness of output is almost as high as that exhibited by my model. Of course, the high negative skewness in that sample is driven almost entirely by the Great Recession.

Simulations of the model imply a zero lower bound (ZLB) incidence between 20 and 30%. Between 1981 and 2018, the U.S. spent 7 out of 48 years, or 14.5% of the time, at the ZLB. ZLB incidence is likely dependent on the natural rate of interest, since a lower equilibrium rate

implies a closer proximity to the constraint. Laubach and Williams (2003) and Laubach and Williams (2016) have estimated the natural real rate to have declined from between 3 and 4% in the 1980s to below 1% since the Great Recession. If this decline proves persistent, the incidence of the ZLB will likely be much higher in the future.

1.6 Micro and Macro Uncertainty

This section develops the paper's main results on micro and macro uncertainty. Macroeconomic uncertainty is an endogenous equilibrium object in this model. Let \mathbf{Z}_t denote the vector of all aggregate variables, such as the wage rate, the capital price, inflation etc., and $Z_t \in \mathbf{Z}_t$. The notion of uncertainty I adopt in this paper is one of conditional forecast errors, as is typical in most of the uncertainty literature. Specifically, I define

$$\mathcal{U}_t(Z_{t+s}) = \sqrt{\frac{1}{s} \mathbb{E}_t \left[\left(Z_{t+s} - \mathbb{E}_t(Z_{t+s}) \right)^2 \right]} \quad (1.24)$$

as uncertainty over future realizations in the aggregate process $\{Z_t\}$. While the macroeconomic uncertainty faced by households extends to the broader set of all components of the stochastic process $\{Z_t\}$, I will oftentimes use uncertainty over economic activity, in short GDP $\{Y_t\}$, as a proxy for macroeconomic uncertainty. And since output follows the process $dY_t = \mu_Y(\Gamma_t)dt + \sigma_Y(\Gamma_t)dB_t$ in equilibrium, I concretely refer to

$$\mathcal{U}_t(Y_{t+dt}) = \underbrace{\sigma_Y(\Gamma_t)}_{\text{"Macroeconomic uncertainty"}}$$

as macro uncertainty at time t .

Isolating the effects of macro uncertainty. The first goal of this section is to quantitatively assess the novel transmission channels identified in Section 1.2. This requires a procedure to isolate and decompose the effects of variation in macro uncertainty. While this is possible for equilibrium objects that explicitly appear in the equilibrium conditions of the model, I am not aware of such a procedure for endogenous uncertainty, which is instead a property of the model's probability

distribution.⁴⁸

To study the transmission mechanism of macro uncertainty in the quantitative model, I instead propose to consider an exogenous fundamental risk shock, $\Delta\sigma_\rho$. This requires adding σ_ρ explicitly to the model's aggregate state space. The assumption implicit in this strategy is that the transmission mechanism of an exogenous increase in the volatility of discount rate shocks, σ_ρ , is sufficiently similar to that of an endogenous increase in the volatility of economic activity, σ_Y . Indeed, I show formally in Section 1.6.3 that endogenous macroeconomic uncertainty is proportional to exogenous fundamental risk, which justifies this approach. Overall, the following analysis of an exogenous fundamental risk shock should be understood primarily as a means to better understand the transmission mechanism of endogenous variation in macroeconomic uncertainty.

Comparison benchmarks. To illustrate my main results, I compare three distinct model benchmarks. The first, which I refer to as the “HANK (full)” benchmark, is the full quantitative model presented in Section 1.3. The second benchmark, which I refer to as “HANK (constant λ^j)”, holds employment transition rates constant and therefore shuts off the interaction between micro and macro uncertainty. Importantly, these two model benchmarks are identical in all other regards. Therefore, a comparison between these two models allows us to approximately isolate the implications of the interaction between micro and macro uncertainty. Finally, I also relate my results to a “RANK” baseline, which is the representative-household New Keynesian model most closely associated with the full HANK model of Section 1.3.

1.6.1 Transmission mechanism of macro uncertainty

As discussed above, I study the transmission mechanism of an exogenous perturbation in fundamental risk, $\Delta\sigma_\rho$, in order to proxy for and better understand the effects of macroeconomic

⁴⁸Kaplan *et al.* (2018) propose a procedure to decompose the partial effects of equilibrium objects that explicitly appear in the household problem, such as the interest rate r_t . Their method applies to the sequence representation of a model without aggregate risk. I generalize their approach to state space representations of models with aggregate risk. However, neither approach allows for a decomposition of the partial effect of endogenous uncertainty. This is because, unlike the aggregate price r_t , uncertainty over its future realizations, σ_r , does not explicitly appear in the household problem and can therefore not “be held fixed”.

uncertainty in this model.⁴⁹ Higher uncertainty elicits an increase in households' desired precautionary savings. This direct effect of uncertainty, borrowing the language of Kaplan *et al.* (2018), leads to an initial contraction in aggregate demand. In the two-period model of Section 1.2, this effect comprises both the direct macro uncertainty channel ① and the direct interaction channels ② and ④. As household consumption falls, firms' demand for labor decreases in general equilibrium. Consequently, the job finding rate, λ_t^U , falls while the separation rate, λ_t^E , rises. Households are now confronted with an endogenous increase in micro uncertainty: The employed are more likely to be laid off and the unemployed are less likely to find a new job. This GE effect that follows the contraction in aggregate demand corresponds to channel ③ in the two-period model.

Figure 1.4 plots both the first- and second-moment response of employment transition rates after an increase in fundamental risk. In expectation, the fundamental risk shock makes it more likely for households to be laid off ($\lambda^E \uparrow$) and less likely to find jobs ($\lambda^U \downarrow$). At the same time, the conditional variance over future realizations of the job finding and separation rates increases substantially. In short, a fundamental risk shock at the macro level translates into uncertainty over job prospects at the micro level.

From the household's perspective, the prospect of job loss is akin to an idiosyncratic rare disaster as I discussed in Section 1.2: It looms large when compared to relatively small and transitory changes in the aggregate wage rate, for example. Endogenous variation in the transition rates into and out of employment has both a first- and second-moment implication for households' future earnings. Consider an employed household. An increase in the job separation rate lowers the expected value of future income because the probability of unemployment rises. At the same time, the variance of earnings increases as well.

The overall effect of a fundamental risk shock is summarized in Figure 1.5. It compares

⁴⁹While my interest in studying the exogenous perturbation $\Delta\sigma_\rho$ is as a proxy for variation in endogenous macro uncertainty, this approach of course builds on a long tradition of studying exogenous second-moment shocks. Following the Great Recession, much work has studied uncertainty as a potential driver of business cycle fluctuations (e.g. Bloom (2009) and Bloom *et al.* (2018) in the context of firms, Bayer *et al.* (2019) focuses on micro uncertainty shocks). Basu and Bundick (2017) is a natural reference point for my paper. They show that, in a representative-agent New Keynesian (RANK) model, a fundamental risk shock generates co-movement in output, consumption, investment and hours worked. In a setting where households have Epstein-Zin preferences with a relative risk aversion coefficient of 80, they find that a 90% increase in the volatility of discount rate shocks generates a 0.2% fall in output due to an increase in desired precautionary savings.

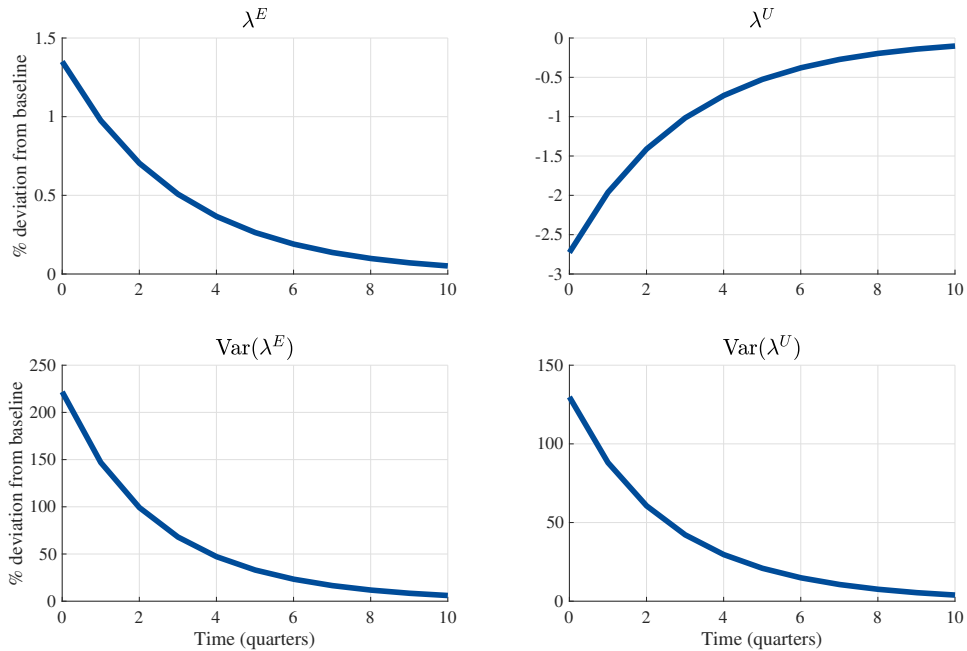


Figure 1.4: Impulse responses of micro uncertainty to $\Delta\sigma_\rho$

Notes. Comparison of impulse responses of employment transition rates to fundamental risk shock. The shock has a half-life of 1 quarter and is initialized at half the size of the Basu and Bundick (2017) shock to facilitate comparison. λ^E and λ^U correspond to the job separation and finding rates, respectively. The bottom row panels plot the response of conditional forecast errors, $\text{Var}_t(\lambda_{t+dt}^j)$.

the impulse responses of output, consumption, investment and hours across the three model benchmarks. As in Basu and Bundick (2017), a fundamental risk shock generates co-movement across the desired aggregates. The main result of this exercise is that the effect of macro uncertainty on aggregate demand is substantially amplified and indeed driven by its interaction with micro uncertainty. The on-impact decline in output in the full model is up to 8 times larger than in the comparison benchmark that holds employment transition rates constant and thereby shuts off any interaction between micro and macro uncertainty. Similarly, the on-impact decline in output in the full model is roughly 5 times larger than in the associated RANK benchmark.⁵⁰

⁵⁰Basu and Bundick (2017) is an instructive reference point for these quantitative results. They study a fundamental risk shock that is about twice as large as the one I consider here and show that, when households have Epstein-Zin preferences with a relative risk aversion coefficient of 80, the on-impact decline in output is about 0.2%. Their model is, of course, different from and in many aspects richer than the RANK baseline I consider here. Nonetheless, it is noteworthy that I reproduce quantitatively similar magnitudes in a setting where households have CRRA preferences

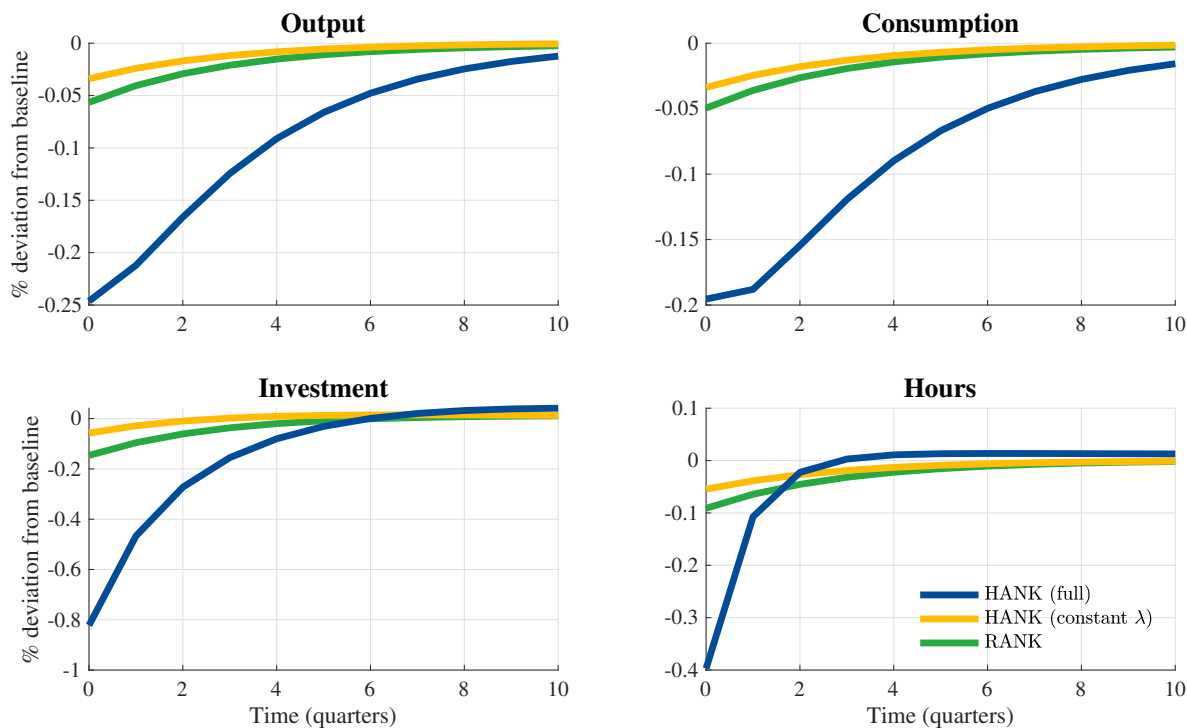


Figure 1.5: *Impulse responses of business cycle aggregates to $\Delta\sigma_\rho$*

Notes. Comparison of impulse responses of business cycle aggregates to fundamental risk shock across three models. The shock has a half-life of 1 quarter and is initialized at half the size of the Basu and Bundick (2017) shock to facilitate comparison. The blue, yellow and green lines correspond, respectively, to the baseline model, the model that shuts off the interaction between micro and macro uncertainty, and the associated RANK model.

Result 1. *The effects of an exogenous increase in fundamental risk, $\Delta\sigma_\rho$, on economic activity are substantially amplified – by a factor of up to 8 – when accounting for the interaction between micro and macro uncertainty.*

Table 1.2 provides a decomposition of the transmission mechanism of a fundamental risk shock. I decompose the shock’s effect on aggregate consumption into a direct uncertainty channel and a set of indirect (GE) channels. This classification extends Kaplan *et al.* (2018) to a setting with aggregate uncertainty. When considering the direct effect of uncertainty, I hold fixed

with a relative risk aversion coefficient of 2: For a given level of risk aversion, households in my setting are more responsive to a change in macro uncertainty precisely because it translates into micro uncertainty.

Table 1.2: *Transmission mechanism of fundamental risk shock $\Delta\sigma_p$*

Contribution to % change in C_0	HANK		RANK
	<i>Normal times</i>	<i>Crisis region</i>	<i>Normal times</i>
Direct effect: uncertainty (micro and macro)	-0.19	-0.22	-0.05
Indirect effect: micro uncertainty	-0.47	-0.51	0.00
Indirect effect: disposable income	0.58	0.66	0.02
Indirect effect: portfolio returns	0.04	-0.12	-0.03
Other effects	-0.16	-0.12	0.01
Total effect (% change in C_0)	-0.2	-0.31	-0.05

Notes. Numbers correspond to on-impact responses, and partial effects add up to the total effect in the last row. The first two columns correspond to the full model; the last column corresponds to the associated RANK model. Columns 1 and 2 compare the partial and total effects when the economy is initialized in the risky steady state (normal times) and on the cusp of the ZLB (crisis region), respectively. The disposable income channel comprises the effects through w_t , τ_t and H_t , and the portfolio returns channel those through r_t , r_t^k and Q_t .

all aggregate variables that would respond in general equilibrium. Households still realize, however, that volatility going forward is heightened. When computing the contribution of indirect channels, I turn back on, one by one, the GE response of the macroeconomic aggregates households care about.

The two rows of Table 1.2 that are most relevant to this discussion are the first (the direct uncertainty effect) and the second (the indirect effect through micro uncertainty). These conceptually correspond, respectively, to the sum of channels ①, ② and ④ (direct) and channel ③ (indirect) in the two-period model of Section 1.2. Table 1.2 demonstrates that the transmission mechanism of macro uncertainty changes completely relative to a representative-agent benchmark. In the RANK model, the direct macro uncertainty effect (channel ①) is the dominant transmission channel.

The transmission mechanism of macro uncertainty in the full HANK model differs in three key regards. First, the direct uncertainty effect on household behavior (the first row of Table 1.2) is larger in absolute terms but muted relative to other transmission channels. This is consistent with the observation from the two-period model that, while ① is the only direct channel operative in the RANK model, the direct interaction effects ② and ④ emerge in the

HANK model. Second, the indirect interaction effect through micro uncertainty (channel ③ in the two-period model, and the second row of Table 1.2) emerges as a dominant transmission channel, especially during times of crisis. Finally, and in the spirit of Kaplan *et al.* (2018), other indirect channels become relatively more important than in the RANK baseline.⁵¹

Result 2. *Accounting for the interaction between uncertainty at the micro and macro level substantially changes the transmission mechanism of the fundamental risk shock $\Delta\sigma_\rho$. Its indirect effect through micro uncertainty becomes a dominant channel in the transmission of macro uncertainty to aggregate consumption spending.*

Studying a fundamental risk shock offers a first glimpse at another major theme of this paper. The behavior of uncertainty changes substantially during times of crisis. The effects of macro uncertainty on economic activity are significantly larger when the economy is close to the ZLB. The dashed lines in Figure 1.5 trace the impulse responses to the same fundamental risk shock as above in an economy that is initialized on the cusp of the ZLB. The shock's on-impact effect on output is 50% larger than during normal times. Similarly, Table 1.2 highlights that the transmission mechanism of macro uncertainty works even more strongly through micro uncertainty when the economy is in a crisis. This finding speaks to substantive interactions between non-linearities at the micro and macro level.⁵²

When the economy is in a crisis state, the relationship between macroeconomic aggregates exhibits a degree of skewness.⁵³ In a crisis, a given negative demand shock realization generates a larger contraction in output than the expansion that would follow a similarly sized positive shock realization. In other words, a zero-mean spread in the distribution of future discount

⁵¹The strong positive response through disposable income is largely driven by profit rebates. In the next iteration of this paper, I plan to address this well-known issue by solving the model with both price and wage stickiness.

⁵²Much previous work has studied the implications of the ZLB constraint on monetary policy in the context of representative-agent New Keynesian (RANK) models (e.g. Christiano *et al.* (2011), Fernández-Villaverde *et al.* (2015a) or Plante *et al.* (2018)). My focus here is primarily on the interaction between non-linearities at the micro and macro level. Indeed, I show first in Section 1.6.2 and later in Section 1.7 that the amplification that derives from the ZLB constraint is itself more pronounced when we account for the interaction between micro and macro uncertainty.

⁵³As is true in U.S. business cycle data, my model generates significant negative skewness and positive kurtosis in output, investment, consumption and hours. See Section 1.5.3.

rate shocks, which is how households perceive the fundamental risk shock, no longer simply leads to a second-moment but also to a first-moment direct effect. In particular, due to negative skewness, this mean-zero spread leads to a fall in economic activity in expectation. The effect of macroeconomic uncertainty on aggregate demand is therefore much larger when the economy is in a crisis state, as showcased in Figure 1.5.

The implications of such skewness at the macro level are not unlike the implications of skewness at the micro level, which I discussed earlier. Indeed, when the economy is in a crisis, this non-linearity at the macro level starts to interact with non-linearity at the micro level. Uncertainty over job prospects at the micro level increases more strongly when the initial fundamental risk shock has a direct first-moment effect at the aggregate level. As a result, the relative importance of the interaction between micro and macro uncertainty becomes even larger in the transmission of fundamental risk shocks.

Result 3. *The behavior of uncertainty changes drastically during economic crises. The effect of a fundamental risk shock on economic activity is 50% larger when the economy is already on the cusp of the ZLB. The interaction with micro uncertainty becomes a relatively even more important channel in the transmission of macro uncertainty.*

The model exhibits a non-linearity at the aggregate level during times of crisis that is nearly impossible to capture in a HANK model without aggregate risk or one that is solved using (lower-order) perturbation methods. Indeed, even higher-order perturbation methods may fail because they still take the economy's normal state as the point of approximation. The macro non-linearity that gives rise to my third result does not manifest while the economy is far away from the crisis region, thus requiring a global solution method.

Of the main results in this paper the ones presented in this subsection are most closely framed with reference to previous work on uncertainty shocks. I will thus conclude by briefly situating my contribution. In the household context, this paper is the first to jointly consider variation in micro and macro uncertainty. Relative to Basu and Bundick (2017), for example, I show that its interaction with unemployment risk emerges as the dominant transmission channel of a macro uncertainty shock. And since micro uncertainty is much more prominent

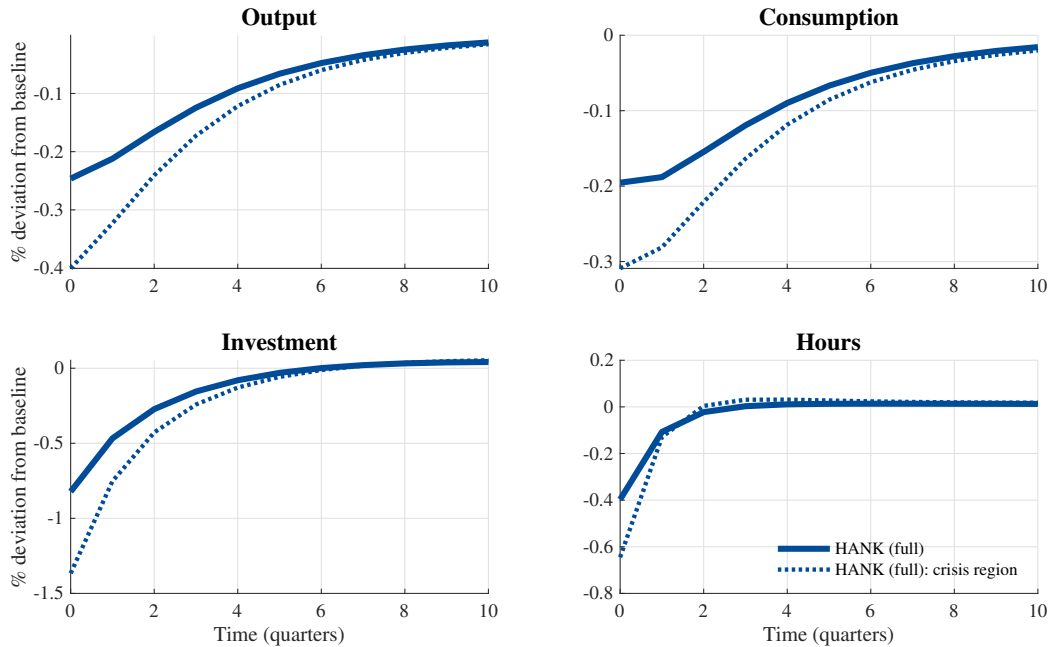


Figure 1.6: Impulse responses of business cycle aggregates to $\Delta\sigma_\rho$

Notes. Comparison of impulse responses of business cycle aggregates to fundamental risk shock across regions of the state space. The shock has a half-life of 1 quarter and is initialized at half the size of the Basu and Bundick (2017) shock to facilitate comparison. The solid lines plot IRFs when the economy is initialized in the risky steady state; the dashed lines plot IRFs when the economy is initialized on the cusp of the ZLB.

from the household’s perspective, my paper matches the magnitude of their quantitative results with a much lower relative risk aversion coefficient. Bayer *et al.* (2019) study micro uncertainty shocks but abstract from time variation in fundamental risk at the aggregate level. They also linearize the macro block of their model and therefore do not capture endogenous variation in macro uncertainty. In the firm context, Bloom *et al.* (2018) model exogenous variation in both the micro and macro uncertainty that firms face. My main insight is that the interaction with *endogenous* micro risk becomes an important transmission channel of macro uncertainty, an interaction that is not modeled in their paper. Lastly, this subsection has demonstrated that the behavior of uncertainty changes in the crisis region of the model, which motivates my emphasis on a global solution method.

1.6.2 Endogenous uncertainty spikes during crises

In this subsection, I show that endogenous variation in uncertainty becomes substantially more pronounced when we account for the interaction between micro and macro uncertainty. In particular, the model generates large endogenous uncertainty spikes in the crisis region. Therefore, a two-way feedback loop emerges between macro uncertainty on the one hand, and micro uncertainty and aggregate demand on the other: Recessions are times when a contraction in economic activity spurs endogenous spikes in uncertainty about the future, which in turn depresses aggregate demand further.

The model features at least two channels through which macroeconomic uncertainty becomes endogenous to economic activity. The first channel centers around the zero lower bound constraint. When the economy is at the ZLB, monetary policy can no longer accommodate negative demand shocks, whose effects on economic activity are consequently amplified. Even when the nominal interest rate is still positive, expansionary monetary policy moves the economy closer to the ZLB, thus raising the likelihood that policy will be constrained in the future. This channel is particularly strong in my baseline model because I do not explicitly take into account the endogeneity of fiscal policy stepping in as a substitute for monetary policy during deep recessions.⁵⁴

The second channel results from the counter-cyclicality in households' average marginal propensity to consume (MPC). As economic activity contracts, households become unemployed and draw down their liquid cash buffers, thus moving closer to their borrowing constraints at the micro level. The prevalence of "hand-to-mouth" behavior grows, which implies an increase in the average household's MPC. Consumer spending, and by implication aggregate activity, thus become more sensitive to further demand shocks. The economy's sensitivity to fundamental risk is precisely the definition of macro uncertainty in my framework, which therefore rises.

This second channel interacts strongly with the fact that the length and severity of a recession are uncertain themselves. Therefore, unlike in settings with no aggregate risk, it becomes difficult for households to properly time the exhaustion of their liquid insurance funds.

⁵⁴Plante *et al.* (2018) show that the ZLB becomes a source of endogenous uncertainty in the context of a representative-agent New Keynesian model.

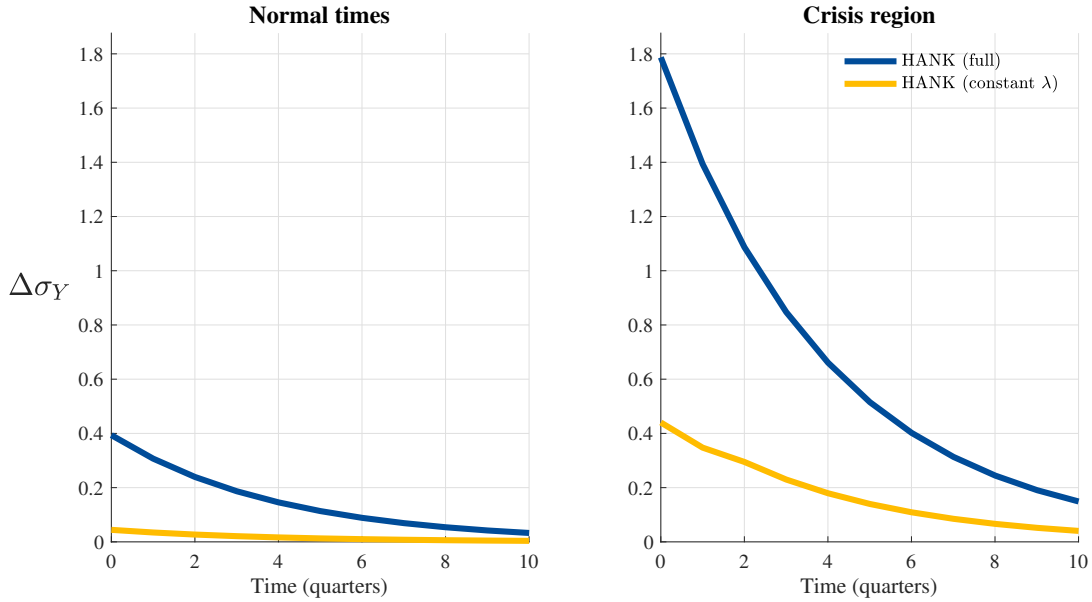


Figure 1.7: Impulse response of endogenous macro uncertainty to $\Delta\rho$

Notes. Comparison of impulse responses of endogenous macro uncertainty, $\sigma_{Y,t}$, to first-moment discount rate shock across different models and state space regions. The shock has a half-life of 1.4 quarters and is initialized at $0.2(\min(\rho_t) - \bar{\rho})$. The blue and yellow lines correspond, respectively, to the baseline model and the model that shuts off the interaction between micro and macro uncertainty. The left and right panels initialize the economy, respectively, in the risky steady state (normal times) and on the cusp of the ZLB (crisis region).

The closer households get to running out of their cash buffers, the more risk averse they become in the face of a potentially prolonged crisis.

What factors determine how sensitive macroeconomic uncertainty is to a given contraction in economic activity? The main novel insight my paper establishes in this context is that first-moment demand shocks have a much larger effect on endogenous macro uncertainty when we account for its interaction with micro uncertainty. Figure 1.7 shows the impulse response of macro uncertainty, $\sigma_Y(\Gamma_t)$, to a negative discount rate shock $\Delta\sigma_\rho$. The left panel shows that the response of endogenous macro uncertainty to a negative demand shock is substantially dampened when we shut off the interaction between micro and macro uncertainty. Indeed in the comparison baseline that holds employment transition rates constant, uncertainty hardly responds at all to the demand shock during normal times. This stark difference further highlights

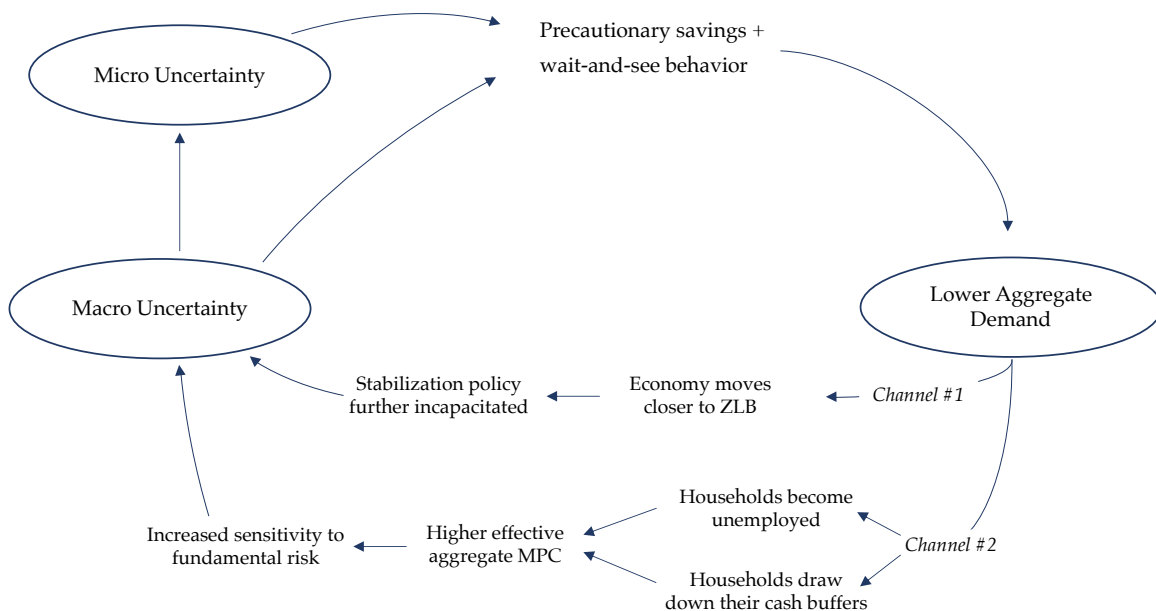


Figure 1.8: GE interaction between micro, macro uncertainty and aggregate demand

why accounting for the interaction between uncertainty at the micro and macro level is crucial in order to understand its role in business cycle fluctuations.

Proximity to the crisis region emerges as another key determinant of the sensitivity of macro uncertainty to economic activity. The right panel traces out the impulse response of macro uncertainty to the same underlying demand shock when the economy starts on the cusp of the crisis region rather than in the risky steady state. The spike in macro uncertainty after a contraction in economic activity during a crisis is over 3 times as large as during normal times. Figure 1.7 also highlights a quantitatively meaningful interaction between non-linearities at the micro and macro level: proximity to the crisis region makes macro uncertainty especially sensitive to economic activity precisely when we account for its interaction with micro uncertainty.

Result 4. *Uncertainty is not only a driver but also an endogenous byproduct of business cycle fluctuations in the full model. The endogenous response in macro uncertainty to a first-moment demand shock is largely driven by the interaction with micro uncertainty. Indeed, uncertainty hardly responds at all to a demand shock during normal times when we shut off the interaction between micro and macro*

uncertainty.

In conclusion, macro uncertainty is most sensitive to aggregate demand and therefore most likely to exhibit endogenous spikes when we account for its interaction with micro uncertainty and when the economy is already in the crisis region.

1.6.3 Uncertainty Multiplier

The discussion thus far highlights that a two-way interaction or feedback loop can emerge in general equilibrium between uncertainty and economic activity. Consistent with the discussion thus far, I illustrate the full feedback loop diagrammatically in Figure 1.8. Uncertainty thus emerges as both a driver and a byproduct of business cycle fluctuations.

I now define an Uncertainty Multiplier, which is a useful tool to characterize the general equilibrium interaction between micro and macro uncertainty. This multiplier measures how much endogenous amplification there is in macroeconomic uncertainty. In particular, it asks how much macroeconomic uncertainty is generated endogenously for a given change in exogenous fundamental risk. The Uncertainty Multiplier is therefore another lens through which we can study the features and implications of uncertainty. It is given by

$$\underbrace{\sigma_Y(\Gamma_t)}_{\text{Macro uncertainty}} = \underbrace{G_Y(\Gamma_t)}_{\text{GE Uncertainty Multiplier}} \cdot \underbrace{\sigma_\rho}_{\text{Exogenous fundamental risk}} \quad (1.25)$$

In particular, this decomposition highlights that macro uncertainty is proportional to exogenous fundamental risk. This justifies my approach in Section 1.6.1 where I studied the fundamental risk shock $\Delta\sigma_\rho$ as a proxy for variation in endogenous macro uncertainty.

In Figure 1.9, I plot the impulse response of endogenous macro uncertainty to an exogenous fundamental risk shock. Consistent with previous discussion, the Uncertainty Multiplier is large when (a) we account for the interaction between micro and macro uncertainty (compare the two models) and (b) when the economy is in the crisis region (for each model, compare the solid to the dashed line). As before, these two forces, non-linearities at the micro and at the macro level, interact.

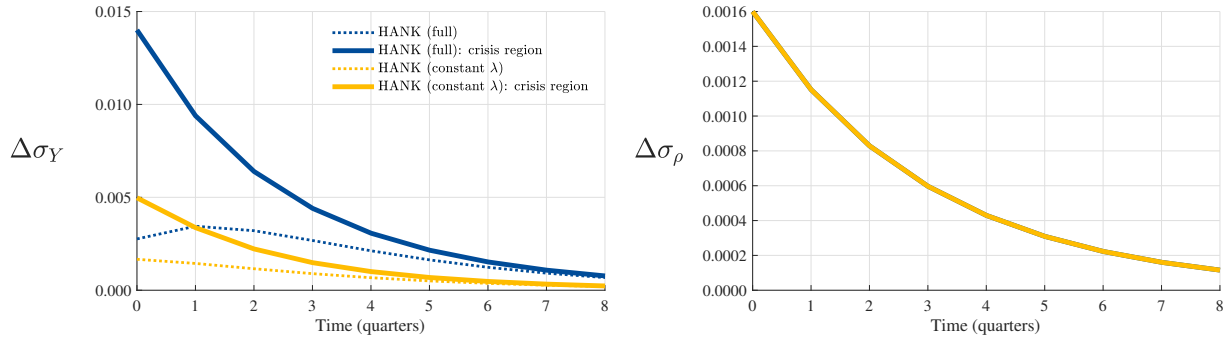


Figure 1.9: *Impulse response of endogenous macro uncertainty to $\Delta\sigma_\rho$*

Notes. Comparison of impulse responses of endogenous macro uncertainty $\sigma_{Y,t}$ to fundamental risk shock across different models and state space regions. The shock, displayed in the right panel, has a half-life of 2 quarters and is initialized at half the size of the Basu and Bundick (2017) shock to facilitate comparison. The blue and yellow lines correspond, respectively, to the baseline model and the model that shuts off the interaction between micro and macro uncertainty. The dashed and solid lines initialize the economy, respectively, in the risky steady state (normal times) and on the cusp of the ZLB (crisis region).

Figure 1.10 offers another perspective on the Uncertainty Multiplier. Instead of plotting IRFs, I offer a state space representation of $G_Y(\Gamma_t)$ (y-axis) against different values of the discount rate shock ρ_t (x-axis). In both models, macroeconomic uncertainty increases endogenously as the economy approaches the crisis region (lower ρ_t). However, Figure 1.10 highlights starkly that the sensitivity of uncertainty to the demand shock is much stronger when we account for the interaction between micro and macro uncertainty in the full model: as we move left in the Figure, the two lines growth increasingly farther apart.

The Uncertainty Multiplier thus offers yet another perspective on the main insight of this paper: there is a two-way feedback loop between macro uncertainty on the one hand, and micro uncertainty and aggregate demand on the other. The model thus generates a natural covariance between uncertainty and economic activity.

1.6.4 Macro uncertainty in the data

This paper proposes a framework in which both micro and macro uncertainty are an endogenous byproduct of recessions but also have strong aggregate demand effects themselves. In general

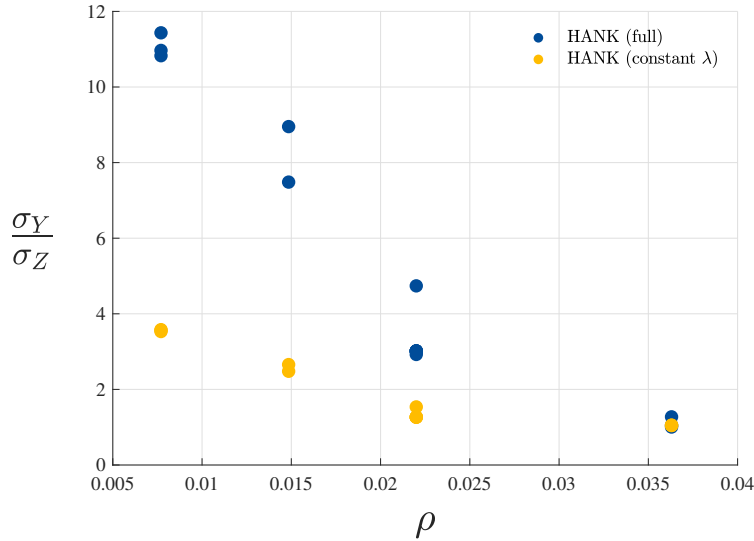


Figure 1.10: State space representation of Uncertainty Multiplier for different ρ

Notes. Comparison of Uncertainty Multiplier $G_Y(\rho, \alpha) = \frac{\sigma_Y(\rho, \alpha)}{\sigma_\rho}$ across different models via state space representation. The x-axis traces out different values in ρ , and the y-axis maps these into $G_Y(\rho, \alpha)$. For a given ρ , vertically aligned dots correspond to different values of α which are added for illustration. The blue and yellow dots correspond, respectively, to the baseline model and the model that shuts off the interaction between micro and macro uncertainty.

equilibrium, economic crises are thus associated with large, endogenous uncertainty spikes, especially when taking into account the interplay between uncertainty at the micro and macro level. This conclusion then naturally begs the question whether the uncertainty generated endogenously in my model is consistent with the data. I tackle this question by comparing the time series moments of macro uncertainty in my model simulations to those of empirical macro uncertainty proxies.

In Table 1.3, I confront the predictions of my model for the cyclicity, persistence, skewness and kurtosis of macro uncertainty with data. These moments are untargeted in my calibration. Furthermore, these simulations only feature first-moment discount rate shocks, and no exogenous second-moment shocks. Empirical work on uncertainty has proposed a range of indices and proxies to measure macro uncertainty in the data. The first row of Table 1.3 corresponds to simulated data from my model and subsequent rows correspond to empirical macro uncertainty

Table 1.3: *Time series moments of macro uncertainty*

Macro uncertainty proxies	Corr w GDP growth	Persistence	Skewness	Kurtosis
Model simulation	-0.33	0.91	0.71	4.14
MacroH1	-0.62	0.92	1.84	7.08
MacroH3	-0.6	0.94	1.85	6.97
MacroH12	-0.6	0.96	1.67	6.49
RealH1	-0.46	0.81	1.21	4.77
FinancialH1	-0.42	0.89	0.88	4.02
VIX	-0.45	0.74	1.94	9.51
Policy Uncertainty	-0.62	0.8	3.13	21.07
World Uncertainty Index	-0.11	0.4	1.58	6.04

Notes. The top row computes time series moments of σ_Y using simulated model data. Subsequent rows report the same time series moments for different empirical macro uncertainty proxies. The indices MacroH*, RealH* and FinancialH* are from Jurado *et al.* (2015). The underlying data can be retrieved from Sydney Ludvigson’s website. The VIX corresponds to the Chicago Board Options Exchange’s Volatility Index. The Policy Uncertainty index is taken from Baker *et al.* (2016) and the World Uncertainty Index is from Ahir *et al.* (2018).

indices. Along the columns, I compare four moments of uncertainty in my model to the data: its cyclical (correlation with GDP growth), persistence, skewness and kurtosis.

The model matches the moments of macro uncertainty in the data surprisingly well: Uncertainty is strongly counter-cyclical, highly persistent, and has high skewness and kurtosis. It is especially noteworthy that I match the high skewness and kurtosis in the data, which bespeaks a degree of non-linearity at the macro level and justifies my emphasis on global solution methods.

Result 5. *Recessions are associated with large, endogenous spikes in uncertainty in the model. Indeed, the time series moments of macroeconomic uncertainty in simulations of the model closely match those in the data: macro uncertainty is strongly counter-cyclical, highly persistent, and features large positive skewness and kurtosis.*

1.7 ZLB Spells and the Paradox of Thrift

Due to the interplay between micro and macro uncertainty, life at the ZLB in my model is even worse than we thought.

1.7.1 ZLB spells with micro and macro uncertainty

The interplay between micro and macro uncertainty makes falling into the ZLB more likely and life at the ZLB more persistent.

In a HANK model with micro and macro uncertainty, at least three channels lead to a higher incidence of ZLB episodes. First, as illustrated in Section 1.6, negative demand shocks are substantially amplified as the economy approaches the ZLB. Second, uncertainty endogenously spikes at both the micro and macro level during economic crises, and an increase in uncertainty has itself a contractionary effect on economic activity. Finally, even during normal times, households anticipate that crisis episodes featuring significant micro and macro risk are more likely. As a result, households perceive the world to be riskier overall, inducing an increase in desired savings and a fall in the natural rate of interest in the risky steady state. The economy is therefore closer to the ZLB constraint even during normal times and monetary policy has less room for counter-cyclical stabilization policy.

1.7.2 Paradox of thrift revisited

“[T]he reactions of the amount of [an individual’s] consumption on the incomes of others makes it impossible for all individuals simultaneously to save any given sums. Every such attempt to save more by reducing consumption will so affect incomes that the attempt necessarily defeats itself. . . [I]t makes [no sense] to neglect [this effect] when we come to aggregate demand.” Keynes (1936), p. 84-85.

The paradox of thrift has long been a central force in macroeconomic theories of business cycle fluctuations. It states that a prudent increase in desired savings at the household level can lead to a contraction of aggregate income in general equilibrium that actually lowers total savings. While the paradox of thrift is an important subject in its own right, its interaction with

the ZLB will be the central focus of this subsection.⁵⁵

What makes the ZLB so destabilizing in many models is that it induces a strong attraction of “self-fulfilling nature” as the economy approaches it. The paradox of thrift is at the heart of this dynamic. Consider the baseline representative-agent New Keynesian model. As the nominal interest rate falls towards zero, agents anticipate that, in future, monetary policy will no longer be able to offset negative shocks with interest rate cuts. Since the ZLB is asymmetric, however, monetary policy will continue to offset large positive shocks with interest rate hikes. As a result, households’ conditional expectation of future real interest rates rises, inducing an increase in desired savings. Since it is not possible for all households to save at the same time, as posited by the paradox of thrift, aggregate income instead falls to offset the rise in aggregate savings demand. This contraction in activity pushes the economy even closer to the ZLB. In any RANK model, this dynamic relies on a strong motive for inter-temporal substitution among households.

The main result of this subsection is to show that, in a HANK economy with micro and macro uncertainty, the general equilibrium attraction exerted by the ZLB is amplified even though the partial equilibrium savings motives across households are dampened. In other words, a different set of channels gives rise to a strong GE interaction between the ZLB and the paradox of thrift, with the main channel operative in a RANK economy substantially dampened.

Figure 1.11 illustrates the amplification generated as the economy approaches the ZLB. The two panels plot output and the nominal interest rate in a state space representation for varying levels of the discount rate shock. As we move left along the x-axis, lowering the discount rate, aggregate demand and thus output fall. As does the nominal interest rate. By definition, the economy reaches the ZLB when the nominal interest rate falls to 0. In the left panel, I indicate this point, the “cusp of the ZLB”, in the state space with a vertical dashed line. Notice that the nominal interest rate falls *linearly* with the discount rate. At a certain point, however, when the economy has come sufficiently close to the ZLB, the “self-fulfilling” amplification effects due to the paradox of thrift kick in as discussed above. In particular, these amplification effects are

⁵⁵Much previous work has studied the interaction between the paradox of thrift and the ZLB in the context of representative-agent or two-agent models. See for example Eggertsson and Woodford (2003), Werning (2011) and Eggertsson and Krugman (2012).

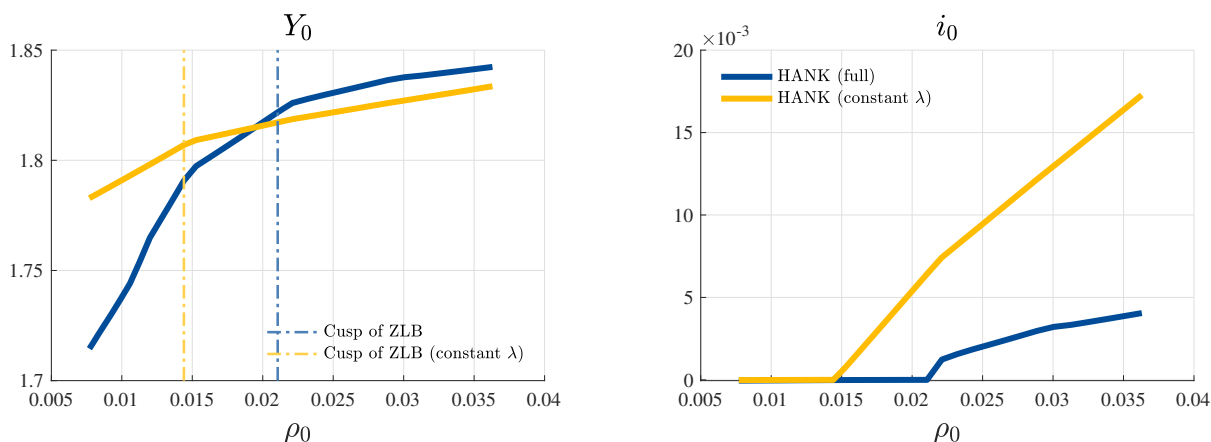


Figure 1.11: Anticipation effects near cusp of ZLB

Notes. Panels trace out on-impact response in output and the nominal interest rate (y-axis) for simulations that initialize the economy at different shock realizations ρ_0 (x-axis). Blue and yellow lines correspond, respectively, to the baseline model and the model that shuts off the interaction between micro and macro uncertainty.

much weaker when we shut off the interaction between micro and macro uncertainty (yellow). In the full model (blue), output also falls linearly with the discount rate initially but becomes increasingly non-linear as the economy first approaches and then falls deeper into the ZLB region.

In the language of Kaplan *et al.* (2018), the set of direct and indirect channels that governs the GE attraction of the ZLB is very different in a HANK model with micro and macro uncertainty. My model not only features a large share of constrained households in normal times; the share of hand-to-mouth households is also strongly counter-cyclical, rising as the economy approaches the ZLB. Liquidity constraints therefore act to dampen the direct effect on desired savings at the micro level because, in the language of Keynes, not all individuals simultaneously try to save.

On the other hand, those households that are on their Euler equations exhibit a stronger rise in desired savings when micro and macro uncertainty interact. Not only do these households anticipate asymmetry in future aggregate states due to the ZLB, as they would in a RANK model, they also internalize the likely spike in unemployment risk. The interplay between micro and macro uncertainty therefore elicits a stronger precautionary motive among unconstrained households.

Finally, as is well known by now, households' consumption and savings decisions in a HANK economy are relatively more influenced by indirect channels in general equilibrium.

Taken together, these factors imply that the interaction between the ZLB and the paradox of thrift is amplified even though not all households increase their desired savings. The overall contraction in aggregate demand posited by the paradox of thrift can be thought of as the product of two effects: a partial equilibrium increase in households' desired savings at the micro level, and a general equilibrium multiplier which maps a given increase in households' precautionary motives into aggregate demand. In a HANK economy with micro and macro uncertainty, the partial equilibrium effect is significantly dampened as many households do not engage in forward-looking consumption smoothing, but the general equilibrium effect is amplified.

1.8 Cost of Business Cycles

The interplay between micro and macro uncertainty has ramifications for the welfare cost of business cycles. In a seminal contribution, Lucas (1987) made the case that the potential gains from stabilization policy are miniscule as he found the welfare cost of business cycle fluctuations, as perceived by a representative household, to be negligible. An active literature has since spawned in response.

The model developed in this paper offers two novel perspectives on this enduring debate. This paper is the first to emphasize that uncertainty spikes endogenously during crises at both the micro and the macro level. Previous work has focused on one or the other.

Second, business cycles are asymmetric. To my knowledge, this is the first paper to compute the cost of business cycles in a globally solved model with a zero lower bound constraint that endogenously generates negative skewness and positive kurtosis in aggregate output, consumption and investment. Beaudry and Pages (2001) and Krebs (2007) compute cost of recession metrics by assuming that, in the context of a model with two aggregate states, stabilization policy can maintain the economy in the good state indefinitely. This paper generates the asymmetry in business cycles that motivates such reasoning endogenously.

To characterize the welfare cost of business cycle fluctuations in my setting, I implement the following thought hypothetical. Consider households that live in the quantitative model of Section 1.3. These households would give up 3.9% of their average consumption to live in the RANK model instead (without ZLB).

1.9 Conclusion

This paper has developed a business cycle model in which micro and macro uncertainty interact. Macro uncertainty shocks are substantially amplified in such a setting, with the indirect effect *through* micro uncertainty emerging as the dominant transmission channel. But macro uncertainty is also endogenous to economic activity, especially during times of crisis. My paper therefore develops a perspective of uncertainty as both a driver and a byproduct of business cycles. I show that accounting for the interaction between uncertainty at the micro and macro level is crucial to understand its broader role in business cycle fluctuations.

Chapter 2

Using Adaptive Sparse Grids to Solve Heterogeneous-Agent Models in Continuous Time¹

2.1 Introduction

Heterogeneous-agent (HA) models continue to proliferate in many fields of economics. Solving dynamic general equilibrium models with rich heterogeneity globally remains a serious challenge, however, due to the curse of dimensionality. HA models oftentimes feature a high-dimensional state space because, in general equilibrium, the cross-sectional distribution of agents becomes part of the aggregate state of the economy. Nevertheless, global solutions are necessary to study many economic applications such as uncertainty, asset pricing, and occasionally binding constraints that give rise to crisis regions. More broadly, many economic phenomena are not well captured by studying linear perturbations around steady states.

The main contribution of this paper is to develop a scalable, robust, and surprisingly portable numerical toolbox to solve high-dimensional partial differential equations (PDEs). In continuous time, a large class of HA models can be represented by (systems of) PDEs. We focus

¹Co-authored with Allen Zhang

on differential equations that arise naturally as the recursive representation of value functions in dynamic programming. Our method leverages adaptive sparse grids to combat the curse of dimensionality, and we develop a consistent and fast solution method for PDEs on irregular grids. Leveraging this method, we compute a global solution of the benchmark Krusell and Smith (1998) model with a 17-dimensional state space in less than 5 minutes on a personal computer. We also solve a continuous-time variant of the Ericson and Pakes (1995) model of dynamic oligopoly and show that, for the same number of grid points, the solution accuracy using our adaptive sparse grid method is an order of magnitude higher than that using standard uniform grids.

Solving general equilibrium models with rich heterogeneity globally is challenging due to the curse of dimensionality. Since the seminal contribution of Smolyak (1963), *sparse grids* have become popular to combat the exponential growth of grid points in higher dimensions. Regular sparse grids are known to reduce the complexity of a grid in d dimensions from $\mathcal{O}(n^d)$ to $\mathcal{O}(n \cdot \log(n)^{d-1})$, where n is the number of unique nodes per dimension (see for example Bungartz and Griebel (2004)). However, further significant efficiency gains can be realized by constructing and refining sparse grids adaptively, placing grid points efficiently in regions of the state space where approximation errors remain large.²

This paper builds on the seminal contribution of Brumm and Scheidegger (2017). They show that adaptive sparse grids are a powerful tool to solve high-dimensional dynamic general equilibrium models in discrete time. Our focus in this paper is on continuous-time applications. The use of continuous time in HA models has proliferated rapidly.³ There are several natural and important computational advantages to the use of continuous time, which are discussed, for example, in Achdou *et al.* (2015).

In this paper, we show that the method of Brumm and Scheidegger (2017) is not readily portable to continuous-time applications. In continuous time, dynamic programming relies on the solution of partial differential equations. The numerical implementation of PDEs on

²See for example Brumm and Scheidegger (2017).

³In macroeconomics, see for example Achdou *et al.* (2015), Ahn *et al.* (2017), Kaplan *et al.* (2018), Fernández-Villaverde *et al.* (2019).

a grid in turn requires the discretization of differential operators. Finite-difference schemes are arguably the most popular such discretization method.⁴ Previous work in the applied mathematics literature has emphasized that the standard finite-difference stencils that lead to consistent discretization schemes on regular grids break down and lead to inconsistent solutions on irregular grids (see for example Schiekofe (1998), Griebel (1998) and Bungartz and Griebel (2004)). As a result, the direct application of adaptive sparse grid methods to solve high-dimensional models in continuous time is not viable.

In this paper, we leverage insights from the applied mathematics literature and make two original contributions that, together, allow us to develop a scalable, robust, and portable continuous-time value function iteration algorithm (or PDE solver) on adaptive sparse grids. The main contribution of this paper is therefore the numerical toolbox accompanying this paper which automates most steps of this algorithm, hopefully allowing for broad accessibility.

Leveraging results in Schiekofe (1998) and Griebel (1998), we develop a generalized *sparse finite-difference method* that leads to consistent solutions of continuous-time heterogeneous-agent models in economics on irregular grids. Our toolbox fully automates this method, providing a function that computes all desired sparse finite-difference operators associated with any given, arbitrarily irregular grid. The efficiency and accessibility of our method are further improved by the two original contributions we make in this paper. First, we show that the differential equations that arise in continuous-time dynamic programming share a common mathematical structure, which we can exploit to streamline the solution of such equations significantly.⁵ In particular, we develop an automated routine with minimal input requirements from the user. Second, we develop a new systematic approach to handle the boundary conditions associated with differential equations. Our toolbox again automates much of this process, thus minimizing user input requirements.

We showcase the broad applicability and scalability of our method by solving high-dimensional benchmark applications from different fields in economics. In Section 2.4, we

⁴Most recent applications in economics employ finite-difference methods. See e.g. Achdou *et al.* (2015).

⁵In their seminal contribution, Achdou *et al.* (2015) also exploit this structure. Building on their work, we develop a routine that computes discretized finite difference operators once ex ante, which improves efficiency and further reduces user input requirements.

solve a continuous-time version of the Krusell and Smith (1998) model using a high-dimensional distribution representation following Schaab (2020). This model has become an important benchmark for evaluating global solutions of heterogeneous-agent models in macroeconomics (see e.g. Den Haan (2010)). Our method computes the global solution of the model and its simulation on a 17-dimensional state space in less than 5 minutes on a personal computer.

In Section 2.5, we solve a modified Ericson and Pakes (1995) model of quality ladders under dynamic oligopolistic competition. Numerous applications in the industrial organization literature have employed variants of this model to study industry dynamics surrounding entry, exit, and investment in settings with strategic firms. Similar to other recent work (e.g. Doraszelski and Judd (2012), Arcidiacono *et al.* (2016), and Blevins (2018)), we cast our model in continuous time. Directly comparing solutions of this model on a regular uniform grid to our method using adaptive sparse grids, we show that adaptive sparse grids achieve roughly one order of magnitude higher accuracy than regular grids for a given number of nodes.

Related literature. Starting with the seminal contributions of Smolyak (1963) and Zenger (1991), sparse grids have enjoyed much popularity in several scientific fields. Krueger and Kubler (2004) and Judd *et al.* (2014) were among the first applications in economics. Our work builds on and is closely related to the implementation in Brumm and Scheidegger (2017) who were the first to employ truly adaptive sparse grid techniques in economics. In this paper, we show that their method is not directly portable to continuous-time applications because the use of standard discretization methods leads to inconsistent solutions of partial differential equations on irregular grids.

We develop a solution method and accompanying numerical toolbox that addresses this shortcoming, allowing for the use of adaptive sparse grid methods to solve continuous-time heterogeneous-agent models in economics. Our method applies and builds on previous results in applied mathematics on the solution of PDEs on irregular grids. Key results on the consistency and stability of finite-difference schemes on irregular sparse grids were developed by Griebel (1998) and Schiekofner (1998).⁶

⁶A partial list of subsequent contributions includes Schiekofner and Zumbusch (1998), Zumbusch (2000), Gerstner

The papers most closely related to ours are Ruttscheidt (2018) and Garcke and Ruttscheidt (2019). While they also use adaptive sparse grids to solve heterogeneous-agent models in continuous time, our paper differs from theirs in three regards. First, their most complex application is a six-dimensional model with two assets and four earnings factors. We demonstrate the scalability of our method by solving applications with many more dimensions. Second, they use interpolation techniques and ghost nodes to guarantee the consistency of their PDE discretization scheme, whereas we use specific basis transformations (as for example in Schiekofe (1998)). Third, we develop a flexible method for handling Dirichlet, von Neumann and other boundary conditions which can be fully automated to minimize user input requirements.⁷

Lastly, our paper contributes to a long literature on global solutions for heterogeneous-agent models.⁸ We contribute to this literature by developing a robust and scalable method for high-dimensional dynamic programming problems in continuous time using adaptive sparse grids. In our applications, we demonstrate the power of our method by globally solving benchmark heterogeneous-agent models from different fields in economics in high dimensions.⁹

2.2 High-Dimensional Partial Differential Equations

In continuous time, a large class of heterogeneous-agent (HA) models can be represented by (systems of) partial differential equations (PDEs). For expositional clarity and illustration, we frame our discussion in Sections 2.2 and 2.3 in terms of a particular PDE, but stress that our methods extend to a large class of (systems of) differential equations. We choose as our

and Griebel (2003), Bungartz and Griebel (2004), Ma and Zabaras (2009), Bokanowski *et al.* (2013), Garcke and Kröner (2017), Ruttscheidt (2018), and Garcke and Ruttscheidt (2019).

⁷Another closely related paper is Ahn (2019). This paper uses a finite volume method to discretize the partial differential equation that characterizes the evolution of cross-sectional distributions in HA models. He shows that finite volume discretization leads to a consistent solution, and guarantees both mass conservation and positivity of the cross-sectional distribution. Our focus, instead, is on finite-difference methods and continuous-time dynamic programming problems that give rise to Hamilton-Jacobi-Bellman equations.

⁸For example, see Den Haan (1996), Den Haan and Others (1997), Krusell and Smith (1998), Reiter (2010), Algan *et al.* (2008), Algan *et al.* (2014), Brunnermeier and Sannikov (2014), Brumm and Scheidegger (2017), Duarte (2018), Pröhl (2019), and Fernández-Villaverde *et al.* (2019).

⁹More recently, grid-free methods have been proposed that leverage deep learning techniques to solve high-dimensional HA models. The relative performance of adaptive sparse grid and deep learning methods in non-trivial applications remains an open question.

leading example a Hamilton-Jacobi-Bellman (HJB) equation which arises as the natural recursive representation of the value function associated with a stochastic control problem. After we present this equation as our leading example, the remainder of this section proceeds in three steps: In Section 2.2.1, we develop the construction of sparse grids, motivated by the curse of dimensionality afflicting dense grids in high dimensions. We discuss two important bases which we use to represent functions on sparse grids in Section 2.2.2. Finally, in Section 2.2.3, we introduce operators that map between these bases and will be crucial in the construction of a consistent finite-difference scheme on sparse grids in Section 2.3. Our discussion of sparse grids in Sections 2.2.1 through 2.2.3 follows closely much previous work on the subject.¹⁰ We again emphasize that our contribution relative to previous work is not in the use of sparse grid methods per se, but rather the general-purpose numerical toolbox we develop for solving high-dimensional heterogeneous-agent models.

We consider the dynamic optimization problem of an agent whose state is given by $\mathbf{X} \in \Omega \subset \mathbb{R}^d$, with $\mathbf{X} = (x_1, \dots, x_d)$ comprising both endogenous and exogenous state variables. The agent's state is assumed to follow the Ito process

$$d\mathbf{X} = \mu(c, \mathbf{X})dt + \sigma(c, \mathbf{X})dB,$$

where $\mu : \mathcal{C} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathcal{C} \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$, and B is m -dimensional Brownian motion. We denote by $c \in \mathcal{C}$ the vector of choice variables under the agent's control which may lie in the admissible set \mathcal{C} .¹¹ The agent's problem is to choose c to maximize the expected discounted sum of future payoff flows which we denote by $u(c)$. It is well known that the value function $V : \Omega \rightarrow \mathbb{R}$ associated with the agent's stochastic control problem is a solution of the HJB equation

$$\rho(\mathbf{X})V(\mathbf{X}) = \max_c \left\{ u(c) + \sum_{i=1}^d \mu_i(c, \mathbf{X})\partial_{x_i}V(\mathbf{X}) + \frac{1}{2} \sum_{i,j=1}^d \sigma_{ij}(c, \mathbf{X})\partial_{x_i x_j}V(\mathbf{X}) \right\}, \quad (2.1)$$

where $\rho : \mathbb{R}^d \rightarrow \mathbb{R}$ is the agent's discount rate, μ_i is the i th element of the vector μ , and, abusing

¹⁰We follow particularly closely recent work by Schiekofe (1998), Brumm and Scheidegger (2017), Ruttscheidt (2018), and Garcke and Ruttscheidt (2019).

¹¹We restrict our attention to stationary Markov controls, so that $c = c(\mathbf{X})$.

notation slightly, σ_{ij} is the ij th element of the matrix $\sigma\sigma^T$. We use the shorthand notation ∂_{x_i} to denote the partial derivative operator $\frac{\partial}{\partial x_i}$, and similarly for $\partial_{x_i x_j}$.¹²

Equation (2.1) describes a stationary or time-independent HJB equation. We focus on economic applications that admit a stationary representation after an appropriate choice of \mathbf{X} . We furthermore assume that equation (2.1) admits a solution with bounded mixed derivatives, and we consider the unit cube as state space, $\Omega = [0, 1]^d$.¹³

Boundary conditions. While equation (2.1) describes the value function V in the interior of the domain, $(0, 1)^d$, boundary conditions must be specified to characterize the behavior of V on the boundary. These boundary conditions naturally vary across economic applications.¹⁴ Mathematically speaking, most boundary conditions encountered in economic applications fall into one of three categories: Dirichlet, von-Neumann, and mixed. The method we present in Section 2.3.3 provides a flexible way to handle these, and we postpone further discussion of boundary conditions until then.

Curse of dimensionality. In many HA applications d is large and equation (2.1) becomes a high-dimensional PDE. In the presence of rich heterogeneity, solving for general equilibrium in dynamic models generally requires tracking the evolution of the cross-sectional distribution of agents as part of the economy's aggregate state. Even coarse approximations of this time-varying distribution can result in a high-dimensional state space when the underlying model does not admit approximate aggregation (Krusell and Smith (1998)). As a result, the curse of dimensionality first discussed by Bellman (1961) remains a serious impediment in practice to solving HA models globally: If n unique nodes are used in each dimension, then the number of grid points required to construct a regular grid in d dimensions is n^d .¹⁵

¹²For the formal derivation of equation (2.1) see, for example, Oksendal (2013) pages 243-251.

¹³This convention is without loss of generality as long as Ω is a rectangular domain. We will provide the appropriate formulas to map between the unit cube and any rectangular domain.

¹⁴In the context of viscosity solutions of HJB equations, we frequently encounter so-called state-constrained boundary conditions. See for example Achdou *et al.* (2015).

¹⁵See for example Bungartz and Griebel (2004) for a detailed discussion of the complexity of solutions uniform grids.

2.2.1 Sparse Grid Construction

Since the seminal contribution of Smolyak (1963), *sparse grids* have become popular to combat the curse of dimensionality when representing functions on high-dimensional grids. Regular sparse grids reduce the complexity of a grid in d dimensions from $\mathcal{O}(n^d)$ to $\mathcal{O}(n \cdot \log(n)^{d-1})$ (see for example Bungartz and Griebel (2004)). Regular sparse grids (or sometimes *Smolyak grids*) have found regular use in economics (see for example Krueger and Kubler (2004) and Judd *et al.* (2014)). However, further significant efficiency gains can be realized by constructing sparse grids adaptively (Brumm and Scheidegger (2017)). In this subsection, we briefly discuss the construction of regular sparse grids, following closely previous work (e.g. Bungartz and Griebel (2004), Brumm and Scheidegger (2017) and Ruttscheidt (2018)).

Dyadic grids. It will be particularly convenient to work with *dyadic grids* (e.g. Brumm and Scheidegger (2017)). A dyadic grid is one whose mesh size in dimension k is defined as 2^{-l_k} for some integer l_k which we call the *level* of the grid in dimension k . A *dense* dyadic grid, then, is one whose grid points in each dimension are distributed equidistantly, each 2^{-l_k} apart from its neighbors along dimension k , such that there are a total of $2^{l_k} - 1$ unique points in dimension k .

Let G be a possibly sparse dyadic grid on Ω . We define the *level of grid* G to be the vector-valued index $l \in \mathbb{N}^d$, with $l = (l_1, \dots, l_d)$. The level index tells us what the finest mesh size of grid G is in each of the d dimensions. For reference, we denote the finest mesh size by $h_l = (h_{l_1}, \dots, h_{l_d}) = (2^{-l_1}, \dots, 2^{-l_d}) \equiv 2^{-l}$.

When constructing grids numerically, we find it convenient to work with the following data structure.¹⁶ Let the vector-valued index i be defined as $i = (i_1, \dots, i_d)$ for $i_k \in \{0, 1, \dots, 2^{l_k}\}$. In d dimensions, any grid point is a tuple of d coordinates. The grid G associated with level index l then comprises the set of *grid points*

$$\left\{ x_{l,i} = (x_{l_1 i_1}, \dots, x_{l_d i_d}) \equiv (i_1 h_{l_1}, \dots, i_d h_{l_d}) \mid 0 \leq i_k \leq 2^{l_k}, i_k \in \mathbb{N} \text{ for all } k \right\}.$$

This definition highlights that i_k denotes the position of the associated grid point in dimension k .

¹⁶See e.g. Schiekofer and Zumbusch (1998) for a broader discussion of efficient data structures for solving PDEs on adaptive sparse grids.

Given the set of grid points, we store our grid using the data structure

$$\mathbf{G} = \begin{pmatrix} (0, \dots, 0) \\ \vdots \\ (x_{l_1, i_1}, \dots, x_{l_d, i_d}) \\ \vdots \\ (1, \dots, 1) \end{pmatrix}$$

Letting J denote the number of grid points, our numerical grid object G is a $J \times d$ matrix.¹⁷

For a given grid G and its associated level index l , we denote by \tilde{G} the *dense dyadic grid* associated with l on the unit cube. A dense grid, given level index l , is a grid that contains the grid points $x_{l,i}$ associated with all possible combinations of $\mathbf{i} \in \mathbb{N}^d$ such that $\mathbf{0} \leq \mathbf{i} \leq 2^l$.

Regular sparse grids. We can call any grid G *sparse* if the set of its grid points is a proper subset of the set of grid points in the dense grid \tilde{G} .¹⁸ Key to the efficient and systematic construction of sparse grids is the observation that any dense grid can be represented as a union of *elementary grids*. Given indices l , any dense grid can be constructed as

$$\tilde{G} = \bigcup_{e \leq l} G_e$$

where elementary grid G_e is defined as

$$G_e = \left\{ (i_1 h_{e_1}, \dots, i_d h_{e_d}) \mid 0 \leq i_k \leq 2^{e_k} \text{ with } i_k \in \mathbb{N} \text{ for all } k, \text{ and } i_k \text{ is odd} \right\}.$$

We illustrate the elementary grids corresponding to levels 1 through 4 in Figure 2.1. Any level

¹⁷We will sometimes abuse notation by referring to G interchangeably as a matrix and a set of grid points (vectors).

¹⁸One way to think of sparse grid G is as a subset of all those index locations that are associated with the dense grid. In particular, let $I_G \subset I_{\tilde{G}}$ denote the set of indices corresponding to grid points in G . That is

$$I_G = \left\{ \mathbf{i} \mid x_{l,i} \in \mathbf{G} \right\}.$$

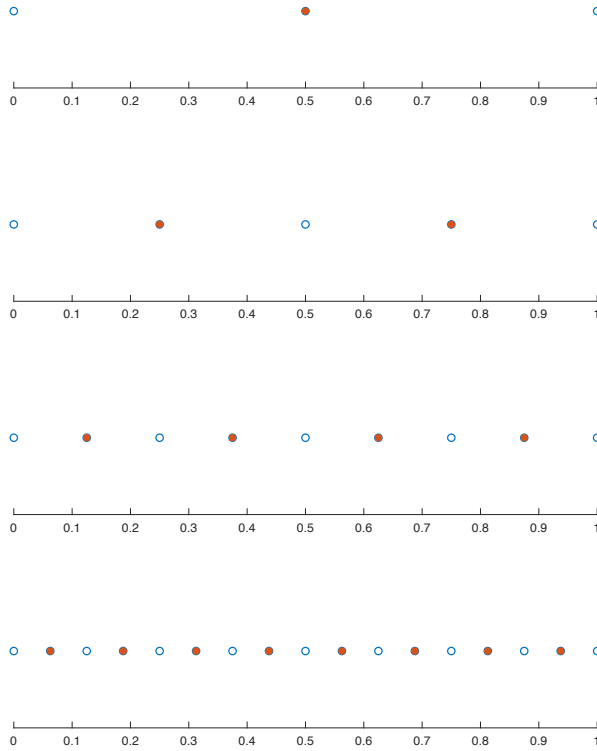


Figure 2.1: *Elementary grids*

index l is associated with a unique dense grid.^{19,20}

We formally construct regular sparse grids by placing restrictions on the union $\cup_{e \leq l} G_e$ which defines the dense grid associated with l . While any restriction would result in a sparse grid, previous work has identified particularly efficient ones.²¹ One such condition is a simple restriction on the sum of elementary grid indices that are used in the union, i.e. $\sum_{k \leq d} e_k = \kappa$ for some constant κ . We illustrate such a restriction on the union of elementary grids in Figure 2.2, where Panel (b) displays the resulting regular sparse grid in two dimensions.

¹⁹For expositional clarity, we start our discussion by focusing on grids with no points on the boundary of the domain. We discuss boundary points below.

²⁰For an illustrative example, consider the case $d = 2$ and the elementary grid for $e = (1, 1)$. We must have $G_{(1,1)} = (0.5, 0.5)$. To see this, we start by constructing the location index i . For $e = (1, 1)$, we have $0 \leq i \leq 2^{(1,1)}$, with the additional requirement that i be odd. The only odd integer between 0 and 2 is 1. Therefore, the elementary grid $G_{(1,1)}$ must be a set of only one point, and this point must be associated with the location index $(1, 1)$. The grid point itself is then given by the formula $x_{e,i} = i \cdot h_e = (1, 1) \cdot (2^{-1}, 2^{-1}) = (\frac{1}{2}, \frac{1}{2})$.

²¹See Bungartz and Griebel (2004) for a discussion.

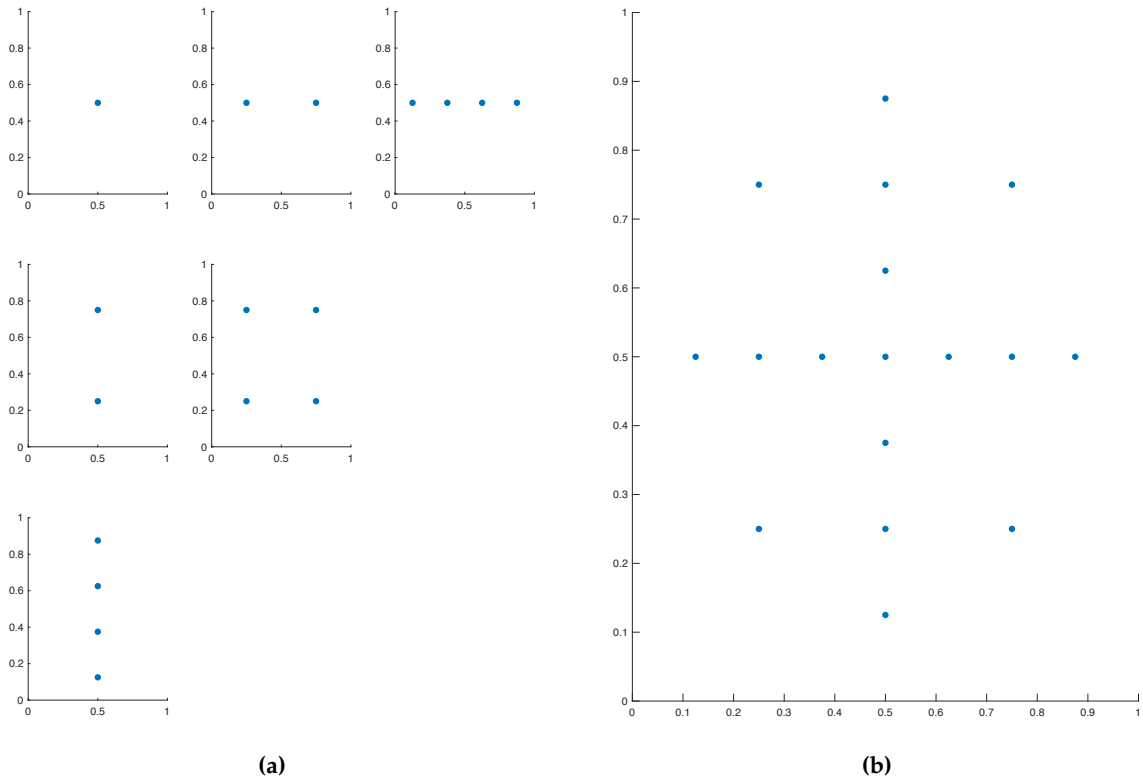


Figure 2.2: Union of elementary grids

Boundary points. We add boundary points in the above construction by considering elementary grids of *level 0*. In practice, we will denote boundary points to be *level 1* and the mid-point currently defined as level 1 to be level 0. This convention saves a lot of grid points as our sparse grids scale to higher dimensions.²² Since the grid points associated with the lowest level (0) must be contained in every grid that adds points associated with higher levels, it is beneficial to associate level 0 with the lowest number of grid points. The particular convention we adopt here is also known as the Clenshaw-Curtis grid (see e.g. Klimke and Wohlmuth (2005)).

Adaptive sparse grids. Any grid G is sparse when it is a proper subset of its dense complement, given a level index l . We distinguish between regular sparse grids, which comprise all sparse grids constructed using a systematic cutoff rule, and adaptive sparse grids, which refine a given grid according to a local approximation error metric. In practice, adaptive sparse grids become

²²See e.g. Brumm and Scheidegger (2017) for a discussion of this.

highly irregular. We discuss adaptation criteria based on equation (2.1) in Section 2.3.4.

2.2.2 Function Representations in the Hierarchical Basis

Consider any grid G that discretizes the domain $\Omega = [0, 1]^d$ on which equation (2.1) is defined. We denote by $V = V(G)$ a discretized grid representation of the value function. In particular, V is a $J \times 1$ vector. It is useful to think of the grid representation V as a combination of basis functions whose value coincides with that of V when evaluated on $G \subset \Omega$. The typical basis used for such representations is the *nodal basis*. In this section, we demonstrate that the nodal basis becomes highly inefficient on adaptive sparse grids, and we introduce the hierarchical basis as a more appropriate alternative. Our subsequent discussion will highlight that the hierarchical basis is the crucial foundation for constructing consistent solutions to differential equation on sparse grids. The advantages of the hierarchical basis have been discussed, for example, in Zenger (1991), Griebel (1998), Brumm and Scheidegger (2017), and Ruttscheidt (2018). Our discussion and notational conventions follow these papers closely.

Nodal basis for dense grids. We construct the nodal basis in terms of *hat functions*. Consider first $d = 1$. Any one-dimensional symmetric hat function can be defined via translation and dilation of the *mother hat function* ϕ as

$$\phi_{l,i}(x) = \phi\left(\frac{x - ih_l}{h_l}\right), \quad \text{where } \phi(x) = \begin{cases} 1 - |x| & \text{for } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

for any set of points $x \in [0, 1]$, where h_l denotes the mesh size of the underlying dense grid. The index pair (l, i) indicates that the hat function is centered around the grid point $x_{l,i}$.

For $d > 1$, we construct higher-dimensional nodal basis functions using the one-dimensional hat functions defined in (2.2). Consider again the grid $G \subset \Omega = [0, 1]^d$ associated with level index l , as well as its dense complement \tilde{G} . For all $x = (x_1, \dots, x_d) \in [0, 1]^d$, we define the d -dimensional nodal basis function centered around the grid point at position \mathbf{i} by

$$\phi_{l,\mathbf{i}}(x) = \prod_{k=1}^d \phi_{l_k, i_k}(x_k). \quad (2.3)$$

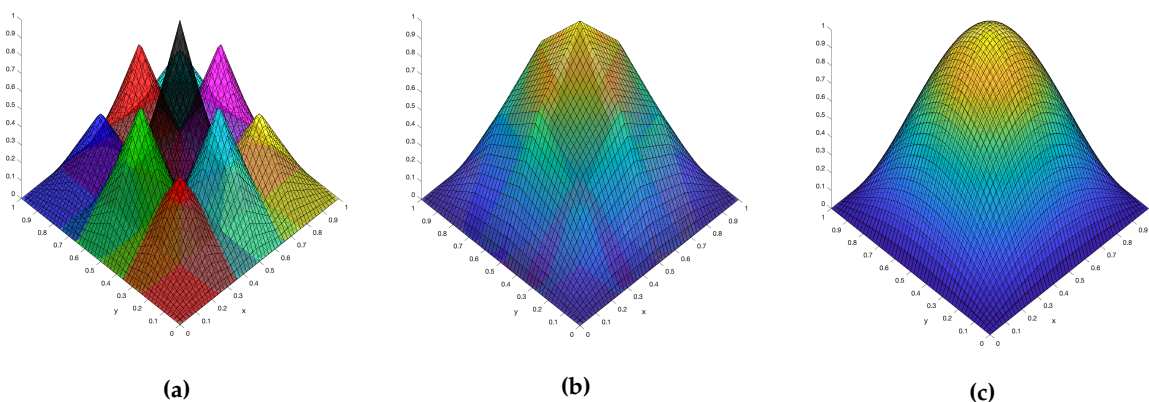


Figure 2.3: *Nodal basis representation of function*

Notes. Nodal basis representation of function defined on $[0, 1]^2$. Panel (a) displays the nodal basis functions in 2D. Panel (b) constructs the interpolation onto the grid $[0, 1]^2$. Panel (c) compares the interpolated approximation to the true function.

By construction, any grid-valued function can be represented in this nodal basis. For example,

$$V = V(\tilde{G}) = \sum_{i \in I_{\tilde{G}}} \alpha_i \phi_{l,i}(\tilde{G})$$

for some *nodal coefficients* α_i . The set of all basis functions $\{\phi_{l,i}(x) \mid i \in I_{\tilde{G}} \text{ and } x \in [0, 1]^d\}$ forms the *nodal basis* in d dimensions.²³ We illustrate the approximation of a two-dimensional function on $[0, 1]^2$ using nodal basis functions in Figure 2.3. Panel (a) displays the nodal basis functions, Panel (b) the function approximation, and Panel (c) the original function.

Nodal basis for sparse grids. On sparse and irregular grids, grid points are no longer equidistantly distributed. As a result, symmetric hat functions can no longer provide a nodal basis for functions represented on such grids. We now introduce *generalized hat functions* and construct an appropriate nodal basis on irregular grids. We again begin with a discussion of the case $d = 1$ and afterwards generalize to higher dimensions.

We define a generalized one-dimensional hat function $\phi_{l,i} : [0, 1] \rightarrow \mathbb{R}$ to be centered at $x_{l,i}$

²³As discussed, for example, in Griebel (1998), Schiekofler (1998) and Ruttscheidt (2018), these piecewise linear nodal basis functions form a finite-dimensional function space, in which we can approximate functions such as V .

with support $[x_{l,i} - h_L, x_{l,i} + h_R] = [ih_l - h_L, ih_l + h_R]$, where the parameter pair (h_L, h_R) denotes the support endpoints. The hat function is then given by

$$\phi_{l,i}(x; c, h_L, h_R) = \begin{cases} 1 - \frac{x-c}{h_R} & \text{for } x \in [c, c + h_R] \\ 1 - \frac{c-x}{h_L} & \text{for } x \in [c - h_L, c] \\ 0 & \text{otherwise} \end{cases} \quad (2.4)$$

for all $x \in [0, 1]$, where $c = x_{l,i}$. It is straightforward to show that this generalized hat function collapses to the simple symmetric hat function defined in (2.2). On a dense grid, $h_L = h_R = h_l$ and the definition (2.4) coincides with (2.2).

When $d > 1$, we again define higher-dimensional generalized hat functions using the tensor product construction in (2.3). As before, the set of generalized hat functions $\{\phi_{l,i}(x) \mid i \in I_G \text{ and } x \in [0, 1]^d\}$ forms a nodal basis for functions represented on the irregular grid G . In particular on irregular grid G , we have $V = V(G) = \sum_{i \in I_G} \alpha_i \phi_{l,i}(G)$. The nodal coefficients are again given by the function evaluations on the grid, that is

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \mathbf{V} = V(G).$$

Drawbacks to the nodal basis. It is now obvious that the d -dimensional nodal basis generally depends on the underlying grid in a quite complicated way. The generalized hat functions used as basis functions are not invariant to the structure of the underlying grid. In practice, therefore, every adaptation of the sparse grid would require constructing a new set of basis functions. This process is not only tedious and potentially costly, but it also fails to preserve valuable information from one iteration of grid adaptation to the next. In Figure 2.4, we illustrate how one-dimensional generalized nodal basis functions change as the underlying regular grid of Panel (a) is adapted in Panel (b).

While the ubiquitous nodal basis is attractive and robust for use on regular grids, it performs poorly on irregular sparse grids. Its inherent drawbacks motivate the search for a

more appropriate basis for work with adaptive sparse grids. We present such a basis in the next subsection.

Hierarchical basis. Representing functions in the hierarchical multi-level basis instead of the common nodal basis is the key step to work efficiently on irregular grids.²⁴

Our goal is to work in a basis that is suitable to sparse grids. In previous sections, we have defined the standard nodal basis function as a symmetric hat function, centered around a grid point at location i and with support everywhere in the interval $[ih_l - h_l, ih_l + h_l]$. These standard basis functions are not compatible with sparse grids because they do not span the *holes* that are characteristic of sparse grids. As we discussed, a natural extension of the standard nodal basis function on a sparse grid is to define the hat function centered around grid point x_i with support everywhere between the neighbor grid points of x_i . Under this construction, the basis functions span all holes. However, the basis functions now depend on the underlying sparse grid and, in particular, the basis functions have to be redefined every time an adjustment to the grid is made.

Working with generalized nodal basis functions can thus become very cumbersome. Hierarchical basis functions, on the other hand, can be defined independently from the underlying grid. Moreover, they span the same function space as the generalized nodal basis functions do for given degrees of freedom.

A set of hierarchical basis functions can be defined as a *restriction* on a corresponding set of standard nodal basis functions. This relationship is the reason for our detailed discussion of the nodal basis. In particular, the set of hierarchical basis functions associated with the levels l and the set of grid point indices I_G is defined as the set of nodal basis functions $\phi_{l,i}(x)$, so that $i \in I_G$ and i is odd. It is this latter requirement which presents the desired restriction on the set of nodal basis functions.

To illustrate, consider the unit interval $[0, 1]$ for $d = 1$. Let \tilde{G} denote the associated dense grid for some level index l . The associated sets of nodal and hierarchical basis functions are then

²⁴See for example Zenger (1991), Griebel (1998) and Bungartz and Griebel (2004), as well as more recently Brumm and Scheidegger (2017) and Ruttsccheidt (2018).

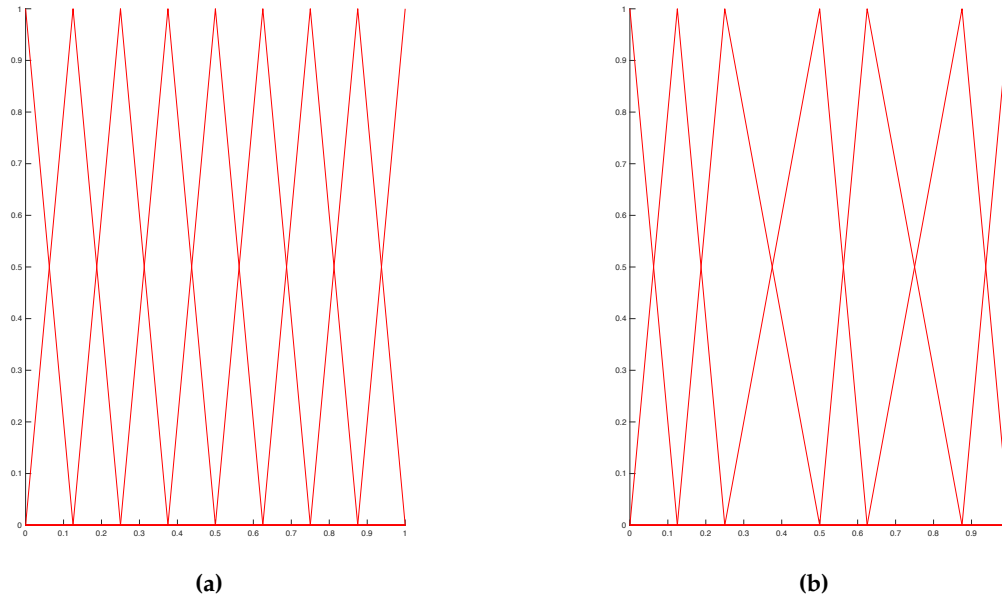


Figure 2.4: *Generalized nodal basis functions*

Notes. Generalized nodal basis functions are no longer invariant with respect to the underlying grid. Panel (a) plots the symmetric hat functions that form a nodal basis on a regular dyadic grid with $l = 5$. Panel (b) demonstrates that, after the grid is adapted and becomes irregular, the nodal basis functions required for constructing a valid basis change.

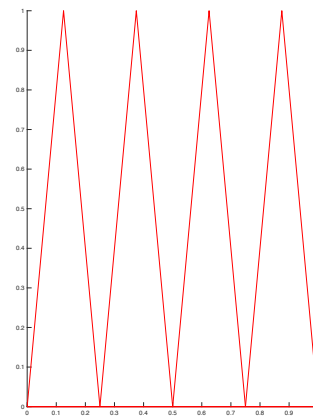
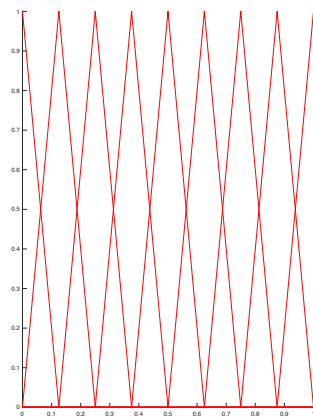
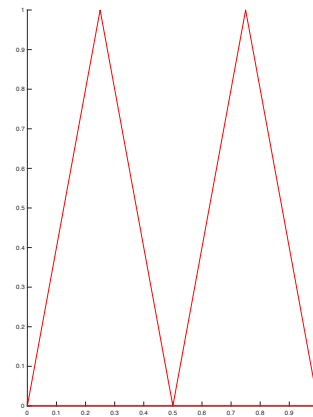
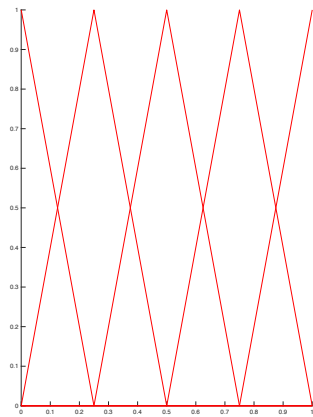
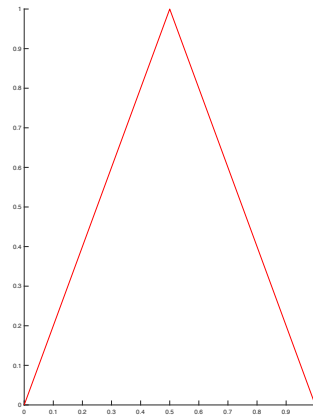
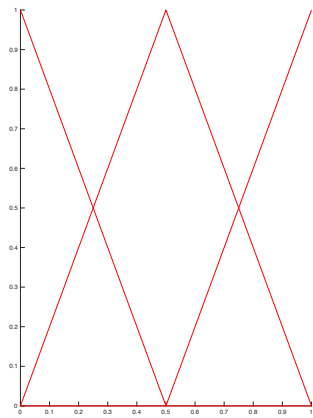
given by²⁵

$$\begin{aligned} \text{Nodal:} & \quad \left\{ \phi_{l,i}(x) \mid 0 \leq i \leq 2^l \text{ and } x \in [0, 1] \right\} \\ \text{Hierarchical:} & \quad \bigcup_{k \leq l} \left\{ \phi_{k,i}(x) \mid 0 \leq i \leq 2^k, \ i \text{ odd, and } x \in [0, 1] \right\}. \end{aligned}$$

The key observation is that any function that is in the span of the nodal basis functions can also be represented by hierarchical basis functions. To do so, however, we need to use the sets of hierarchical basis functions associated with each level below and equal to l . In Figure 2.5, we compare one-dimensional nodal (Panel a) and hierarchical (Panel b) basis functions for levels 1 through 3.

To formalize the relationship between the nodal basis and the hierarchical basis, we adopt

²⁵See e.g. Schiekofe (1998).



(a) Nodal basis functions

(b) Hierarchical basis functions

Figure 2.5: Nodal and hierarchical basis functions for different grids

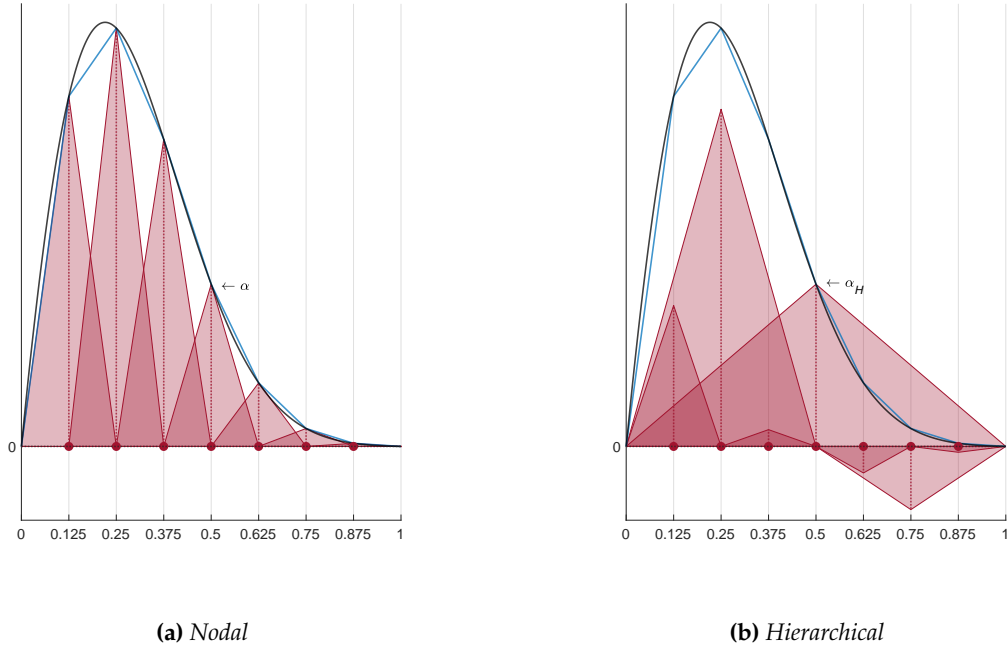


Figure 2.6: Function representation in nodal and hierarchical basis

Notes. The black solid line is the function $f(x) = 1000\phi(3(x + 0.2)) \sin(\pi x)$, where $\phi(x)$ is the Gaussian density function, on $[0, 1]$. Panel (a) approximates $f(x)$ on a regular grid with $l = 3$ in the nodal basis. Panel (b) approximates $f(x)$ on the same regular grid in the hierarchical basis.

the notation from and follow the discussion of Bungartz and Griebel (2004). Denote the space of all functions that can be expressed as linear combinations of standard nodal basis functions as

$$\mathcal{V}_l = \text{span} \left\{ \phi_{l,i}(x) \mid \mathbf{0} \leq i \leq 2^l \text{ and } x \in [0, 1]^d \right\}.$$

And denote the space of all functions that can be expressed as linear combinations of the hierarchical basis functions associated with levels l as

$$\mathcal{W}_l = \text{span} \left\{ \phi_{l,i}(x) \mid \mathbf{0} \leq i \leq 2^l, \ i_k \text{ odd for all } 1 \leq k \leq d, \text{ and } x \in [0, 1]^d \right\}.$$

Then we have

$$\mathcal{V}_l = \bigoplus_{k \leq l} \mathcal{W}_k.$$

In particular, denoting hierarchical basis functions by $\phi_{l,i}^H$ for distinction, if $f \in \mathcal{V}_l$ then

there exists a set of hierarchical coefficients $\{\alpha_{k,i}^H\}$ such that

$$f(x) = \sum_{k \leq l} \sum_{0 \leq i < 2^k} \alpha_{k,i}^H \phi_{l,i}^H(x) \equiv \sum_{k \leq l} T_k^H(x) \alpha_k^H,$$

where α_k^H is the column vector of hierarchical coefficients and $T_k^H(x)$ is the matrix of hierarchical basis functions of level k .

Representing functions in the hierarchical basis on adapted sparse grids. We claimed at the outset of this subsection that the hierarchical basis is particularly suitable when working with irregular grids. We now work through an illustrative example. Figures 2.6 and 2.7 display approximations of the function $f(x) = 1000\phi(3(x + 0.2)) \sin(\pi x)$, where $\phi(x)$ is the Gaussian density function, on $[0, 1]$. In Figure 2.6, we show this function approximation on a uniform grid with $l = 3$ using nodal basis functions, panel (a), and hierarchical basis functions, panel (b).

Figure 2.7 illustrates that the hierarchical basis is especially convenient when working on irregular grids. Figure 2.7 shows a function approximation of $f(x)$ on an adapted grid. The figure highlights that the hierarchical coefficients corresponding to those basis functions associated with the remaining grid points are unchanged after grid adaptation. See Brumm and Scheidegger (2017) for a similar discussion.

2.2.3 Hierarchization Operators

A representation of the value function $V(X)$ on a grid G is associated with the coefficient vectors α and α^H in the nodal and hierarchical bases, respectively. We now show that there are convenient mappings between these representations.²⁶ Our notation and discussion follow especially closely previous work by Schiekofer (1998), Griebel (1998) and Rutzscheidt (2018).

In one dimension, we denote these operators as

$$H : \alpha \mapsto \alpha^H \quad \text{and} \quad E : \alpha^H \mapsto \alpha,$$

where H is known as the *hierarchization operator* and E as the *de-hierarchization operator*. Our

²⁶These mappings have been previously discussed by Griebel (1998), Schiekofer (1998), Bungartz and Griebel (2004) and Rutzscheidt (2018) among others.

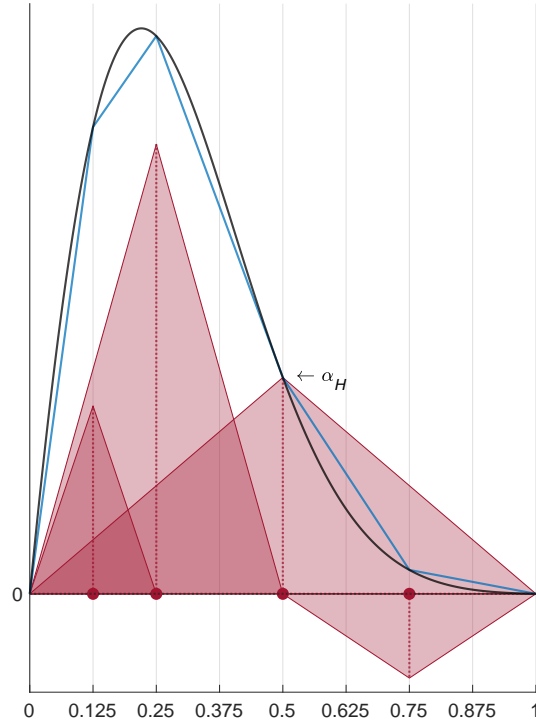


Figure 2.7: Function representation on adapted sparse grid in the hierarchical basis

construction of \mathbf{H} and \mathbf{E} follows Schiekofer (1998).²⁷ In particular, define the hierarchization operator $H_{l,i}^k$, which maps the nodal coefficient associated with grid point x_{l_k,i_k} in dimension k into its associated hierarchical coefficient, by the stencil $[-\frac{1}{2} \quad 1 \quad -\frac{1}{2}]$, applied at point x_{l_k,i_k} and with step size h_k . That is,

$$H_{l,i}^k f(x_{l_k,i_k}) = f(x_{l_k,i_k}) - \frac{f(x_{l_k,i_k} - h_k) + f(x_{l_k,i_k} + h_k)}{2}.$$

The d -dimensional construction follows analogously (see Schiekofer (1998) for details). We now define \mathbf{H}^k as the operator that applies the one-dimensional hierarchization $H_{l,i}^k$ in dimension k

²⁷For a reference in English see, for example, Ruttscheidt (2018).

on all grid points (again, see Schiekofer (1998) or Ruttscheidt (2018)). Then we have

$$\mathbf{H} = \prod_{k=1}^d \mathbf{H}^k.$$

The de-hierarchization operator is simply defined as the inverse mapping, that is $\mathbf{E} = \mathbf{H}^{-1}$.

Importantly, we also introduce the operators

$$\mathbf{H}_k = \prod_{j=1, j \neq k}^d \mathbf{H}^j$$

and $\mathbf{E}_k = \mathbf{H}_k^{-1}$ as performing the hierarchization and de-hierarchization operations in all dimensions *but* dimension k . (See again Schiekofer (1998) or Ruttscheidt (2018)). These operators will be crucial to define consistent finite-difference stencils on irregular grids.

2.3 Value Function Iteration on Adaptive Sparse Grids

We solve equation (2.1) on irregular grid G using a continuous-time value function iteration (VFI) algorithm.

Formally, we employ a *time-marching scheme*, a common strategy to solve stationary PDEs (e.g. LeVeque (2007)).²⁸ A time-marching scheme constructs a sequence of approximate solutions $\{\mathbf{V}^n\}$, given some initial guess \mathbf{V}^0 , as if the problem were iterated forward through time. Convergence to a fixed point with $\mathbf{V}^{n+1} = \mathbf{V}^n$ then implies a solution to the stationary problem.

For a grid-based numerical implementation, equation (2.1) must be discretized. While Section 2.2 proposed an efficient representation of the function V on irregular grid G using the hierarchical basis, in this section we discuss the discretization of the differential operators that appear in equation (2.1). Our discussion of PDE solutions on sparse grids, in particular in Section 2.3.1, again follows closely previous work by e.g. Griebel (1998), Schiekofer (1998), Bungartz and Griebel (2004) and Ruttscheidt (2018).

The search for appropriate discretization schemes of partial differential equations is guided by the objectives of stability and consistency (see e.g. LeVeque (2007) for a discussion). Together, they imply that the discretized approximation \mathbf{V} on $G \subset \Omega$ converges to the true solution $V(X)$,

²⁸In this paper, we exclusively focus on grid-based solution methods for differential equations.

for $X \in \Omega$, as the underlying grid becomes increasingly fine.

Our focus in this paper is on *finite-difference* (FD) methods, arguably the most prevalent solution strategy in practice. Using this approach, we discretize the operators ∂_{x_i} and $\partial_{x_i x_j}$ in equation (2.1) using *finite-difference stencils*.

Consider the *standard finite-difference stencils* for first derivatives associated with dimensions i and j

$$D_i^F = \begin{bmatrix} 0 & -1 & 1 \end{bmatrix} \quad D_i^B = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} \quad D_i^C = \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \quad (2.5)$$

as well as the standard stencils for second-derivative operators

$$D_{ii} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \quad D_{ij} = D_i^C D_j^C \quad (2.6)$$

where D_i^F , D_i^B and D_i^C denote the forward, backward and central stencils for the first-derivative, D_{ii} the second-derivative stencil, and D_{ij} the mixed-derivative stencil in dimensions i and j .

On the dense grid \tilde{G} , a discretization of the differential operators using the standard FD stencils leads to a consistent scheme to solve equation (2.1). See for example Barles and Souganidis (1991) or, more recently, Achdou *et al.* (2015). The value function iteration or time-marching scheme on dense grid \tilde{G} , using the regular FD operators, can then be represented as

$$\frac{\mathbf{V}^{n+1} - \mathbf{V}^n}{\Delta} + \rho \mathbf{V}^{n+\iota} = \mathbf{u}^n + \sum_{i=1}^d \mu_i^n D_i^{\text{upwind}} \mathbf{V}^{n+\iota} + \frac{1}{2} \sum_{i,j=1}^d \sigma_{ij}^n D_{ij} \mathbf{V}^{n+\iota} \quad (2.7)$$

where bold-faced notation is here used for objects discretized on grid \tilde{G} , and where $\iota \in \{0, 1\}$ depending on whether an *explicit* ($\iota = 0$) or *implicit* ($\iota = 1$) scheme is employed. We use bold-faced letters here to represent discretized vectors and matrices on a given grid. Equation (2.7) is a linear system that can be solved for \mathbf{V}^{n+1} in each iteration n of the time-marching scheme.

Two additional observations about equation (2.7) are in order. First, the max operator of equation (2.1) no longer appears here. Formally, the max operator implies a first-order condition that can be solved numerically for the optimal controls or policy functions c^n as a function of the current guess \mathbf{V}^n . The values of c^n are then implicitly subsumed in equation (2.7) in the terms $\mathbf{u}^n = u(c^n)$, $\mu_i = \mu_i(c^n, \tilde{X})$, and $\sigma_{ij} = \sigma_{ij}(c^n, \tilde{X})$. Second, the first-order differential operators ∂_{x_i} are discretized using D_i^{upwind} , an operator we use to denote an implicit *upwind scheme* (see e.g.

Achdou *et al.* (2015)).

It is well known that the discretization scheme (2.7) is consistent on dense grids (Barles and Souganidis (1991), Achdou *et al.* (2015)). If this result generalized to irregular grids such as G , implying that we could represent V on the irregular grid G and construct a VFI algorithm to solve equation (2.1) using the regular stencils defined in (2.5) and (2.6), we would be done. However, a discretization scheme based on the standard FD stencils leads to an inconsistent solution of equation (2.1). For a proof, see for example Schiekofer (1998).

2.3.1 Consistent Finite-Difference Schemes on Adaptive Sparse Grids

The standard FD stencils (2.5) and (2.6) can be generalized to non-uniform (but regular) grids in order to address the uneven spacing of grid points in a given dimension. Intuitively, the standard stencils in dimensions i and j only draw information from nodes in dimensions i and j . That is, these stencils look for direct neighbors in their own dimensions. On highly irregular grids, however, directly neighboring nodes in a given dimension might be far away while a much closer neighbor may be available in a different dimension. We now show that for sufficiently smooth functions, drawing information from closer off-dimension neighbors leads to more accurate and indeed consistent discretization schemes for equation (2.1) on the irregular grid G .²⁹

Consistent finite difference operators on sparse grids can be constructed by leveraging the matrices \mathbf{H} and \mathbf{E} introduced in Section 2.2.3. Our notation and discussion again follow closely that in Schiekofer (1998) and Ruttscheidt (2018). We define the first- and second-derivative *sparse finite-difference operators* in dimension i as

$$D_i^{\text{S,F}} = \mathbf{E}_i D_i^{\text{F}} \mathbf{H}_i \quad D_i^{\text{S,B}} = \mathbf{E}_i D_i^{\text{B}} \mathbf{H}_i \quad D_{ii}^{\text{S}} = \mathbf{E}_i D_{ii} \mathbf{H}_i. \quad (2.8)$$

Since the regular FD operators are pre-multiplied by the de-hierarchization operator \mathbf{E}_i and post-multiplied by the hierarchization operator \mathbf{H}_i , the sparse FD operators must consequently be applied to functions in their nodal representation. We furthermore define the mixed-derivative

²⁹For this reason, sparse grid methods are especially powerful in higher dimensions, where sparse finite difference operators can draw information from many off-dimension neighbors.

sparse FD operator in dimensions i and j as

$$D_{ij}^S = \mathbf{E}_i D_i \mathbf{H}_i \mathbf{E}_j D_j \mathbf{H}_j = D_i^S D_j^S \quad (2.9)$$

Consider now the following discretized time-marching scheme on irregular grid G ,

$$\frac{\mathbf{V}^{n+1} - \mathbf{V}^n}{\Delta} + \rho \mathbf{V}^{n+\iota} = \mathbf{u}^n + \sum_{i=1}^d \mu_i^n D_i^{\text{S,upwind}} \mathbf{V}^{n+\iota} + \frac{1}{2} \sum_{i,j=1}^d \sigma_{ij}^n D_{ij}^S \mathbf{V}^{n+\iota} \quad (2.10)$$

where, as in (2.7), $\iota \in \{0, 1\}$ depending on whether an explicit or implicit scheme is employed. Similarly, the optimal policies c^n are again subsumed. Unlike in (2.7), however, scheme (2.10) now uses the sparse finite-difference operators defined in (2.8) and (2.9). We again use the operator $D_i^{\text{S,upwind}}$ to denote an implicit upwinding scheme involving both $D_i^{\text{S,F}}$ and $D_i^{\text{S,B}}$.

Schiekofer (1998) first showed that the sparse FD operators defined in (2.8) and (2.9) lead to a consistent discretization scheme of partial differential equations such as (2.1) on regular sparse grids. In particular, he shows that, under some regularity conditions on V , the sparse FD operators lead to consistent discretizations of the associated partial derivative operators.³⁰ As a result, the discretization scheme (2.10) leads to a consistent solution of (2.1) on sparse grid G . When sparse grids are adaptively refined, the regular finite-difference stencils associated with non-equidistant grids must be used (see e.g. Griebel (1998)).

2.3.2 Efficient Operator Construction

In practice, many papers that implement VFI schemes such as (2.7) construct the associated discretized FD operators manually. That approach not only limits portability but also implies that large matrices are reconstructed with every iteration n of the VFI.

We now propose an alternative approach, exploiting the mathematical structure of Hamilton-Jacobi-Bellman equations such as (2.1). Our method is based on two observations. First, the terms $\mu_i^n D_i^{\text{S,upwind}}$ and $\sigma_{ij}^n D_{ij}^S$ do not have to be recomputed from scratch in each iteration n . In particular, given any grid G , all sparse FD operators can be pre-computed as matrices prior to the VFI step, with the sole exception of certain boundary conditions which we address below

³⁰In particular, V must have bounded mixed derivatives.

in Section 2.3.3. Consequently, only μ_i^n and σ_{ij}^n are recomputed in iteration n , allowing for a significantly more efficient construction of the terms of the linear system (2.10).

Second, the construction of linear system (2.10) can be further streamlined by rewriting it as

$$\begin{aligned} \frac{\mathbf{V}^{n+1} - \mathbf{V}^n}{\Delta} + \rho \mathbf{V}^{n+l} &= \mathbf{u}^n + \left(\sum_{i=1}^d \mu_i^n D_i^{\text{S,upwind}} + \frac{1}{2} \sum_{i,j=1}^d \sigma_{ij}^n D_{ij}^{\text{S}} \right) \mathbf{V}^{n+l} \\ &= \mathbf{u}^n + \mathbf{A}^n \mathbf{V}^{n+l}. \end{aligned}$$

In particular, since the construction of the FD matrices is frontloaded, we can implement a function that constructs the matrix \mathbf{A}^n automatically given inputs μ_i^n and σ_{ij}^n . Formally,

$$\mathbf{A}^n = \varphi(\mu_i^n, \sigma_{ij}^n, \mathbf{G}). \quad (2.11)$$

Evaluating the function φ is not particularly costly. In practice, therefore, the only burden on the user is to provide a function that computes μ_i^n and σ_{ij}^n given a grid \mathbf{G} and a value function guess \mathbf{V}^n in iteration n . Given this mapping, which evidently depends on the underlying economic application, our toolbox provides automated routines to directly solve for \mathbf{V}^{n+1} without further user input.

2.3.3 Boundary Conditions

To complete the definition of partial differential equation (2.1) and, by extension, the VFI scheme (2.10) a set of boundary conditions must be specified.³¹ Although these boundary conditions vary with the economic application under consideration, we can discuss their mathematical structure and develop a robust numerical implementation.

We discuss the two most prevalent types of boundary conditions in this setting, Dirichlet and Von Neumann. Because of differences in information contained in each, the two require different treatments. For each type of boundary condition, we demonstrate our treatment via the following one-dimensional working example.

³¹Since our focus in this paper is on stationary problems, we focus on boundary conditions rather than initial-value problems.

Consider a particular problem nested by equation (2.1) and defined on $\Omega = [0, 1]$. We denote the boundary of the state space by $\partial\Omega = \{0, 1\}$. The HJB equation holds on the interior the domain $(0, 1)$. We discretize a one-dimensional grid associated with indices $\{1, \dots, J\}$, where $x_1 = 0$ and $x_J = 1$. We denote by V_i element of the discretized V associated with grid point i . We discuss them in the context of linear system (2.10) which, assuming an implicit method with $\iota = 1$, can be rewritten in the spirit of Section 2.3.2 as

$$\mathbf{B}^n \mathbf{V}^{n+1} = \mathbf{b}^n, \quad (2.12)$$

where

$$\mathbf{B}^n = \left(\frac{1}{\Delta} + \boldsymbol{\rho} - \mathbf{A}^n \right), \quad \text{and } \mathbf{b}^n = \mathbf{u}^n + \frac{1}{\Delta} \mathbf{V}^n.$$

Dirichlet. A set of Dirichlet boundary conditions consists of $V|_{\partial\Omega} = \kappa$, which says that the *level* of V along the entire boundary is specified. In the one-dimensional problem this specifies that $V(0) = \kappa_L$ and $V(1) = \kappa_R$ for given scalars κ_L and κ_R . Here the linear system (2.12) still holds for the interior of the domain, i.e. the $J - 2$ points associated with $i \in \{2, \dots, J - 1\}$. But now there are an additional set of constraints that set $V_1^n = V_1^{n+1} = \kappa_L$ and $V_J^n = V_J^{n+1} = \kappa_R$.

The linear system for the interior of the domain is captured in the equation

$$\mathbf{B}_{2:J-1, 2:J-1}^n \mathbf{V}_{2:J-1}^{n+1} = \mathbf{b}_{2:J-1}^n.$$

However, an adjustment is necessary to implement the additional constraints from the boundary condition. The effects of V_1^{n+1} and V_J^{n+1} on this system are captured by the vectors $\mathbf{B}_{1, 2:J-1}^n V_1^{n+1}$ and $\mathbf{B}_{J, 2:J-1}^n V_J^{n+1}$ respectively. Hence, we instead solve

$$\mathbf{B}_{2:J-1, 2:J-1}^n \mathbf{V}_{2:J-1}^{n+1} + \mathbf{B}_{1, 2:J-1}^n \kappa_L + \mathbf{B}_{J, 2:J-1}^n \kappa_R = \mathbf{b}_{2:J-1}^n.$$

to obtain the desired solution.³²

³²Note this is equivalent to solving (2.12) after replacing the first and last rows of the matrix \mathbf{B}^n with an identity matrix (that is, the first row is entirely 0 except for a 1 in the first column, and the last row is entirely 0 except for a 1 in the last column), and the first and last elements of \mathbf{b}^n with κ_L and κ_R respectively.

Von Neumann. Consider the same setting as above, but now with the set of boundary conditions

$$V_x(0) = \kappa_L(V(0)) \quad \text{and} \quad V_x(1) = \kappa_R(V(1)),$$

where we also allow the exogenously given BC to depend on the function itself.

Our implementation here differs from the Dirichlet case. Depending on the desired stencil at the boundary, the derivative at the boundary may not involve any interior points. Hence a solution that only applies the PDE (and thus the linear system) in the interior of the discretized state space may not have any information with which to pin down the boundary values.

To alleviate these concerns, we instead solve the linear system (2.12) on an extended grid by introducing ghost nodes x_0 and x_{J+1} and associated function values V_0 and V_{J+1} . We assume that equation (2.1) also applies to the actual boundary points of the grid, x_1 and x_J , and use the ghost nodes to support evaluation of derivatives at these points.

Values at the ghost nodes can be derived from the boundary conditions. Supposing the condition uses central stencils on the boundary (extension to other stencils is straightforward), we have

$$\frac{V_2^n - V_0^n}{2dx} = \kappa_L(V_1^n) \tag{2.13}$$

$$\frac{V_{J+1}^n - V_{J-1}^n}{2dx} = \kappa_R(V_J^n). \tag{2.14}$$

We can now solve for the ghost nodes as

$$V_0^n = V_2^n - 2\kappa_L(V_1^n)dx \tag{2.15}$$

$$V_{J+1}^n = V_{J-1}^n + 2\kappa_R(V_J^n)dx. \tag{2.16}$$

With expressions for V_0^n and V_{J+1}^n in hand, we are now ready to modify equation (2.12) for the solution. As we have shown in equations (2.13) and (2.14), finite difference stencils on the extended grid can place weight on the ghost nodes. Let \tilde{A}^n denote the matrix that is constructed from collecting these weights, pulling them through the expressions in equations (2.15) and (2.16), and applying the construction in equation (2.11). Then we can solve the

following modification of equation (2.12)

$$\mathbf{B}^n \mathbf{V}^{n+1} - \tilde{\mathbf{A}}^n \mathbf{V}^{n+1} = \mathbf{b}^n. \quad (2.17)$$

Intuitively, whenever one of the equations in the linear system requires a value from a ghost node, we reexpress this value in terms of nodes on the original grid and add its influence into the system. This implementation ensures that the boundary conditions are satisfied at every iteration.

2.3.4 Grid Adaptation

Sections 2.3.1 through 2.3.3 discuss how to solve $\mathbf{V} \approx V(\mathbf{G})$ on a given sparse grid \mathbf{G} . The power and efficiency of adaptive sparse grid methods lies in the iterative adaptation of grid \mathbf{G} , placing new grid points efficiently in regions where the approximation \mathbf{V} is poor while taking away grid points that do not add much to overall accuracy.

We introduce grid adaptation as an iterative loop outside the VFI block. Iteration k of this loop is associated with the grid \mathbf{G}^k and solves an approximate solution to equation (2.1) which we denote \mathbf{V}^k . The goal of constructing an adapted grid \mathbf{G}^{k+1} is to add grid points where the approximation error $\|\mathbf{V} - \mathbf{V}^k\|$ remains relatively large. Formally, the grid \mathbf{G}^k of iteration k induces the function space W^k of piecewise linear functions. The discretized value function approximation \mathbf{V}^k should be thought of as an element of this function space. This is the sense in which we can consider a norm $\|\mathbf{V} - \mathbf{V}^k\|$. Mechanically, the remaining error is large in regions of the domain Ω where V features a lot of concavity or, in other words, where the true V is particularly different from the piecewise linear basis functions operative in that region.

We again leverage the hierarchical basis representation of \mathbf{V}^k to determine an efficient adaptation \mathbf{G}^{k+1} . Due to its multi-level structure, the size of the lowest-level hierarchical coefficient is a good measure of the function's concavity in the region. An adaptation criterion can therefore be chosen simply in terms of the hierarchical coefficients obtained in iteration k . Since nodal coefficients are simply given by $\alpha^k = \mathbf{V}^k$, we have $\alpha^{H,k} = \mathbf{H}^k \mathbf{V}^k$, where $\alpha^{H,k}$ is the $J^k \times 1$ vector of hierarchical surpluses and J^k denotes the number of grid points in grid \mathbf{G}^k . Following Brumm and Scheidegger (2017) (and references therein), we therefore employ the

adaptation criterion

$$\text{add the children of node } 1 \leq i \leq J^k \text{ if: } |\alpha_i^{H,k}| > \epsilon^{\text{add}} \quad (2.18)$$

for some appropriate choice of $0 < \epsilon^{\text{add}}$. Similarly, we employ an adaptation criterion for the removal of nodes as

$$\text{keep node } 1 \leq i \leq J^k \text{ if: } |\alpha_i^{H,k}| > \epsilon^{\text{keep}}, \quad (2.19)$$

for an appropriate choice of $0 < \epsilon^{\text{keep}} < \epsilon^{\text{add}}$.

Whenever a particular node i in the grid satisfies the adaptation criterion, we add all the grid points *one level lower* than node i . We denote these grid points the *children* of the *parent* node i . During the adaptation process, it is crucial that the grid contains all the parents of all of its nodes. If this is not the case and the grid has *holes*, we are no longer able to map nodal into hierarchical function representations (see e.g. Ruttscheidt (2018) for a discussion). In practice, we check to add any parent nodes back in if our adaptation algorithm deletes them.

2.3.5 Algorithm Structure

Our proposed algorithm to solve equation (2.1) is summarized in Algorithm 1 below.

Algorithm 1 Value Function Iteration on Adaptive Sparse Grids

- 1: Choose an index I^0 and construct associated regular sparse grid G^0
 - 2: Construct H^0 and sparse FD operators on G^0 ▷ Equations (2.8) and (2.9)
 - 3: Initialize guess for the value function V^0 on G^0
 - 4: **for** $k \geq 0$ **do**
 - 5: **for** $n \geq 0$ **do**
 - 6: Compute $A^{k,n}$ using sparse FD operators ▷ Equation (2.11)
 - 7: Solve linear system for $V^{k,n+1}$ ▷ Equation (2.12)
 - 8: Continue if $|V^{k,n+1} - V^{k,n}| > \epsilon^{\text{VFI}}$
 - 9: **end**
 - 10: Compute hierarchical surplus $\alpha^{H,k} = H^k V^k$
 - 11: If adaptation criterion met, adapt grid G^{k+1} ▷ Equations (2.18) and (2.19)
 - 12: Reconstruct H^{k+1} and sparse FD operators on G^{k+1} ▷ Equations (2.8) and (2.9)
 - 13: **end**
-

2.4 Krusell-Smith

In this section, we leverage our adaptive sparse grid method to solve the first of our applications, the Krusell and Smith (1998) model. This model has become the benchmark application for testing global solution methods for heterogeneous-agent models. The continuous-time rendition of the model, which we present in Section 2.4.1, follows Ahn *et al.* (2017). In Section 2.4.2, we discuss a high-dimensional distribution approximation following Schaab (2020). Lastly, in Section 2.4.3, we evaluate the performance of our algorithm. We show that we can solve the baseline Krusell-Smith model with a distribution representation that uses 14 dimensions on a personal computer in less than five minutes.

2.4.1 The Krusell and Smith (1998) Model

The model is populated by a continuum of heterogeneous households of mass 1. Households solve a standard consumption-savings problem while facing both idiosyncratic risk in the form of unemployment spells and aggregate risk. While markets are incomplete, households can trade capital which they rent to firms.

Households. Household preferences are given by

$$\mathbb{E}_0 \int_0^{\infty} e^{-\rho t} u(c_t) dt,$$

where ρ is the household discount rate and c_t denotes consumption flow. Households' budget constraint is given by

$$\dot{k}_t = r_t k_t + (1 - \tau) w_t z_t + \tau_t^{\text{lump}} - c_t,$$

where k_t is the stock of capital owned and

$$r_t = r_t^k - \delta$$

is the rate of return on capital net of depreciation δ , where r_t^k is the rental rate. Household earnings comprise a salary component, where τ is the income tax rate, w_t the real wage rate and z_t the household's idiosyncratic earnings potential, as well as a lump-sum rebate τ_t^{lump}

from the government. We assume that z_t follows a Poisson process with discrete earnings states, discussed below in Section 2.4.3.

Firms. A representative and perfectly competitive firm uses capital and labor to produce the final consumption good according to the aggregate production function

$$Y_t = e^{Z_t} K_t^\alpha N_t^{1-\alpha},$$

where Z_t is aggregate TFP, and K_t and N_t denote the aggregate capital stock and labor, respectively. α denotes the income share of capital.

Firm profits are given by $Y_t - r_t^k K_t - w_t N_t$. With Cobb-Douglas production and under perfect competition, factor prices are then simply given by

$$r_t^k = \alpha \frac{Y_t}{K_t}$$

$$w_t = (1 - \alpha) \frac{Y_t}{L_t}.$$

Government. The role of the fiscal authority in this setting is limited. We abstract from outright government purchases. In fact, the only role of the government in this model is the provision of unemployment insurance, which is funded by a tax on labor. In other words, we assume that

$$\tau_t^{\text{lump}} = \tau^{\text{UI}}(z_t),$$

where the dependence of $\tau^{\text{UI}}(z_t)$ on state z_t allows for the interpretation as conditional unemployment insurance. The government budget constraint imposes a balanced budget, which implies that labor tax revenue must be equal to total UI payments. We solve residually for the income tax rate τ so that the government budget is balanced.

Aggregation and market clearing. We denote by $g_t(k, z)$ the cross-sectional distribution of households in this economy over capital k and employment z . Given a realized cross-sectional

g_t , aggregation at time t is then simply defined, for example for consumption, as

$$\begin{aligned} C_t &= \int \int c_t(k, z) g_t(k, z) dk dz \\ &= \int c_t(k, z^E) g_t(k, z^E) dk + \int c_t(k, z^U) g_t(k, z^U) dk, \end{aligned}$$

where C_t is aggregate consumption. The second equation makes use of our assumption that $z_t \in \{z^E, z^U\}$ follows a two-state Markov chain. The aggregate capital stock is then similarly given by $K_t = \int \int k g_t(k, z) dk dz$, and aggregate labor by $N_t = \int \int z g_t(k, z) dk dz$ since we assume that all households inelastically supply one unit of labor.

There are three markets in this economy, for goods, capital and labor. The markets for capital and labor clear implicitly by imposing the above aggregation relationships. They imply, respectively, that capital rented by the representative firm, K_t , must be equal to the capital owned by all households, and likewise for labor. The market for the consumption good then clears automatically by Walras' law.

Recursive problem and state space. Aggregate TFP Z_t is the only aggregate risk factor of the model. We assume it follows a mean-reverting process, with

$$dZ_t = -\theta Z_t dt + \sigma dB_t,$$

where B_t is standard Brownian motion.

The aggregate state of our economy is then given by (Z_t, g_t) . Since households must forecast future prices, which in turn depend on future realizations of the cross-sectional distribution, households must also forecast the entire cross-sectional distribution g_t , which consequently enters the state space. The state space of the household's dynamic problem is then given by (k_t, z_t, Z_t, g_t) .

In particular, the Hamilton-Jacobi-Bellman equation associated with a recursive formulation

of the household problem is given by

$$\begin{aligned} \rho V(k, z, Z, g) = \max_c \left\{ u(c) + V_k(k, z, Z, g) \left[r(Z, g)k + (1 - \tau)w(Z, g)z + \tau^{\text{UI}}(z) - c \right] \right\} \quad (2.20) \\ + \lambda_{z,z'} \left[V(k, z', Z, g) - V(k, z, Z, g) \right] - \theta Z V_Z(k, z, Z, g) + \frac{\sigma^2}{2} V_{ZZ}(k, z, Z, g) \\ + \int \frac{\delta V(k, z, Z, g)}{\delta g(k, z)} (\mathcal{A}^* g)(k, z) dk dz, \end{aligned}$$

where $\lambda_{z,z'}$ denotes the continuous-time hazard rate of transitioning from employment state z to z' . The third line in equation (2.20) captures the effect of expected changes in the cross-sectional distribution on today's value function (see for example Schaab (2020) and Ahn *et al.* (2017)). \mathcal{A}^* is the operator associated with the Kolmogorov forward equation that characterizes the evolution of the cross-sectional distribution. In particular,

$$\begin{aligned} dg(k, z) = (\mathcal{A}^* g)(k, z) dt = - \partial_k \left[\left(r(Z, g)k + (1 - \tau)w(Z, g)z + \tau^{\text{UI}}(z) - c(k, z, Z, g) \right) \right] \quad (2.21) \\ - \lambda_{z,z'} g(k, z) + \lambda_{z',z} g(k, z'). \end{aligned}$$

2.4.2 High-Dimensional Distribution Approximation

The HJB equation (2.20) does not yet fall into the framework presented in Sections 2.2 and 2.3 because it is defined on the infinite-dimensional state space (k, z, Z, g) . To solve the Krusell-Smith model, an approximation assumption is needed. To showcase our adaptive sparse grid algorithm in the context of a high-dimensional application, we follow Schaab (2020) and posit a high- but finite-dimensional approximation of the cross-sectional distribution, given by

$$g_t(k, z) \approx \hat{g}_t(k, z) = F(\boldsymbol{\alpha}_t)(k, z), \quad (2.22)$$

where F is a set of basis functions that are parametrized by $\boldsymbol{\alpha}_t \in \mathbb{R}^N$. As in Schaab (2020), we solve an *approximate economy* that takes \hat{g}_t as its cross-sectional distribution rather than g_t . With an increasingly accurate representation $g_t(k, z) \approx F(\boldsymbol{\alpha}_t)(k, z)$, which we achieve by increasing N for an appropriate class of basis functions F , the approximate economy converges to the true model.

Under the approximation assumption (2.22), the recursive household problem associated

with the approximate economy is given by

$$\begin{aligned} \rho \hat{V}(k, z, Z, \alpha) = \max_c \left\{ u(c) + \hat{V}_k(k, z, Z, \alpha) \left[\hat{r}(Z, \alpha)k + (1 - \tau)\hat{w}(Z, \alpha)z + \tau^{\text{UI}}(z) - c \right] \right\} \quad (2.23) \\ + \lambda_{z,z'} \left[\hat{V}(k, z', Z, \alpha) - \hat{V}(k, z, Z, \alpha) \right] - \theta Z \hat{V}_Z(k, z, Z, \alpha) + \frac{\sigma^2}{2} \hat{V}_{ZZ}(k, z, Z, \alpha) \\ + \hat{V}_\alpha(k, z, Z, \alpha) \mu_\alpha(Z, \alpha), \end{aligned}$$

where the internally consistent law of motion for α_t is given by $d\alpha = \mu_\alpha dt$. For a proof and an analytical characterization of μ_α see Schaab (2020).

The HJB equation (2.23) of the approximate economy is now a special case of equation (2.1), and therefore falls squarely into the framework we have presented. We next discuss our numerical implementation of this approximate economy for particular choices of $F(\cdot)$ using adaptive sparse grids.

2.4.3 Numerical Results

Before evaluating the performance and accuracy of our algorithm, we discuss the calibration we use in Section 2.4.3 and specify what assumptions we make on $F(\cdot)$ in equation (2.22) in Section 2.4.3.

Calibration

On the household side, we calibrate $\rho = 0.05$ and assume CRRA preferences with relative risk aversion of $\gamma = 2$. Households face uninsurable earnings risk encoded in the two-state Markov process $z_t \in \{z^E, z^U\}$. We model z_t using a Poisson process with arrival rates $\lambda(z)$. For simplicity, we set $z^E = 1.2$ and $z^U = 0.8$, as well as $\lambda(z^E) = \lambda(z^U) = 1/3$. While this calibration does not correspond to the interpretation of z_t as employment transitions, it has no significant implications for our focus on comparing solutions of the model for different assumptions on $F(\cdot)$. Similarly, we set $\tau = \tau^{\text{UI}}(z) = 0$.

For aggregate TFP risk, we set $\theta = 0.05$ and $\sigma = 0.007$. On the production, we calibrate $\alpha = 0.33$. And we assume a rate of capital depreciation of $\delta = 0.05$.

Basis Functions $F(\cdot)$

For the main analysis, we use Chebyshev polynomials to approximate the cross-sectional distribution. That is, we assume

$$g_t(k, z) \approx \mathbf{T}(k, z) \boldsymbol{\alpha}_t = \sum_{n=1}^N \alpha_t^n T^n(k, z),$$

where N is again the number of basis functions and $T^n(k, z)$ is the n th Chebyshev polynomial associated with time-varying coefficient α_t^n , which we stack together in $\mathbf{T}(k, z)$ and $\boldsymbol{\alpha}_t$, respectively.

Since z_t is a discrete two-state process, we split up the distribution representation as follows. Let $g_t^E(k) = g_t(k, z^E)$ denote the marginal distribution of capital ownership among households with earnings state z^E , and likewise for z^U . In practice, we then use

$$g_t^E(k) \approx \mathbf{T}(k) \boldsymbol{\alpha}_t^E = \sum_{n=1}^{N^E} \alpha_t^{E,n} T^n(k, z)$$
$$g_t^U(k) \approx \mathbf{T}(k) \boldsymbol{\alpha}_t^U = \sum_{n=1}^{N^U} \alpha_t^{U,n} T^n(k, z),$$

with $N = N^E + N^U$ denoting the total number of basis functions used. For the main analysis, we set $N^E = N^U = 7$. Consequently, we solve the household value function (2.23) of the approximate economy over the 17-dimensional state space $(k, z, Z, \boldsymbol{\alpha}^E, \boldsymbol{\alpha}^U)$.³³

Performance and Accuracy

To assess the accuracy of our solution method, we compute the Den Haan (2010) metric of our model solution, which has become the benchmark accuracy metric used in the literature on global solution methods for heterogeneous-agent models. This error metric compares two simulations of the model for the same draw of aggregate shocks $\{Z_t\}$. The first simulation directly uses the law of motion $d\boldsymbol{\alpha}_t = \boldsymbol{\mu}_\alpha dt$ to simulate the approximate distribution. We denote simulated objects, such as capital, under this simulation by K_t^{lom} .

³³In practice, we can scale up the number of basis functions and, consequently, dimensions substantially beyond this baseline. We choose these values for N^E and N^U so that we can solve the entire model in less than 5 minutes on a personal computer.

The second simulation instead updates the cross-sectional household distribution at each time step by directly simulating the consumption and savings decisions of households at the micro level. In discrete time, the Young (2010) algorithm is typically used to update the distribution. In continuous time, we can directly use the Kolmogorov forward equation to perform this update (see Schaab (2020) for details). We denote simulated objects under this simulation by K_t^{sim} .

The Den Haan (2010) metric is then defined as

$$\epsilon_K^{\text{DH}} = 100 \times \max_t \log \left\{ K_t^{\text{lom}} - K_t^{\text{sim}} \right\}.$$

We can interpret ϵ_K^{DH} as the largest percentage error in an unconditional forecast of capital given the realized draw of shocks $\{Z_t\}$. When ϵ_K^{DH} is small, household beliefs are, by construction, consistent with the true law of motion of the economy, which is the required condition for rational expectations equilibrium.

In Figure 2.8 we plot the simulated time series for K_t^{sim} (solid blue line) and K_t^{lom} (dashed yellow line). The difference between these two lines can therefore be interpreted as (proportional to) the Den Haan (2010) error metric. The figure illustrates that our method achieves an accurate fit as the two lines lie nearly on top of each other. Formally, this simulation of the model achieves an error metric of $\epsilon_K^{\text{DH}} = 0.152$, which implies that the largest unconditional household forecast error is 0.15%. The entire solution and simulation of the model takes less than 5 minutes on a personal computer.³⁴

It is important to emphasize at this point that the experiment presented here is not set up to minimize ϵ_K^{DH} , but rather to highlight our ability to scale up the distribution representation $F(\cdot)$ to high dimensions. For example, our distribution approximation reconstructs the cross-sectional distribution from scratch using Chebyshev polynomials. A particularly straightforward approach to substantially improve accuracy would be to approximate $g_t - g^0$ using polynomials instead of g_t , where g^0 is the distribution associated with the deterministic steady state. We currently do not leverage the information encoded in g^0 at all.

³⁴We use a 2019 16-inch MacBook Pro, with a 2.4 GHz 8-Core Intel i9 processor and 32 GB memory.

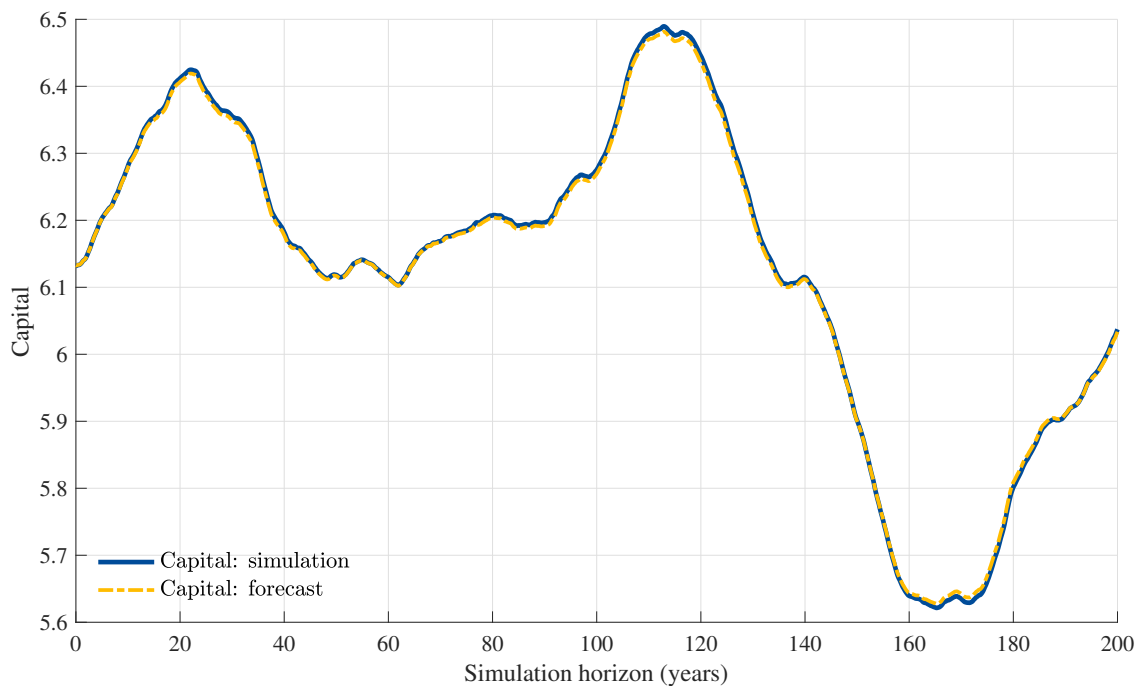


Figure 2.8: Den Haan (2010) metric for solution of the Krusell-Smith model

Notes. The solid blue line corresponds to K_t^{sim} and the dashed yellow line to K_t^{lom} . The difference between these two lines is therefore proportional to the error metric ϵ_K^{DH} .

2.5 Quality Ladders Under Dynamic Oligopoly

For our next application, we study a modified Ericson and Pakes (1995) model of quality ladders under dynamic oligopolistic competition. Similar models have been used extensively in the industrial organization literature to study industry dynamics surrounding entry, exit, and investment in settings with strategic firms. While estimation of such models have been greatly simplified via conditional choice probability estimators (Hotz and Miller (1993), Hotz *et al.* (1994), Bajari *et al.* (2007), and Pakes *et al.* (2007)), equilibrium computation for counterfactuals remains an active area of research.

Recently, Doraszelski and Judd (2012), Arcidiacono *et al.* (2016), Blevins (2018), and others have demonstrated that casting such models in continuous time alleviates some of the computa-

tional burden of previous discrete time formulations of the model.³⁵ We also cast the model in continuous time, with similarities in setup and notation to Blevins (2018). However, we depart from Blevins (2018) by assuming a continuous state variable and reverting back to the continuous control found in Ericson and Pakes (1995). We also abstract away from entry and exit decisions and focus purely on investment.

To demonstrate the effectiveness of our adaptive sparse grid methodology, we first solve the model on a sufficiently large dense grid to use as a benchmark. We then solve the model on smaller grids (both sparse and dense) of varying size, and compare the error rates of the resulting solutions against the benchmark. While our method is designed for maximum computational gains in high-dimensional cases of the model (i.e., many firms), we focus on the case with only two firms in this application to facilitate a comparison against a sufficiently dense benchmark solution.

2.5.1 Model

In the model, N firms each make a single product of quality $\omega_i \in [0, \bar{\omega}]$. Firms compete in the product market, where market shares depend both on the quality of all products available on the market and their accompanying prices. Quality is naturally decreasing over time (representing competition from external markets as in Blevins (2018)), but can also be increased via a continuous investment choice. The aggregate state vector is $\omega \equiv (\omega_1, \dots, \omega_N)$.

A unit mass of consumers derive utility each period from purchasing at most one good. If a consumer purchases the good from firm i , she derives utility equal to $g(\omega_i) - p_i + \epsilon_i$, where ϵ_i is a logit error and $g(\omega_i) = \delta_\pi \log(\omega_i + 1)$. The outside option (i.e., no purchase) is normalized to $\omega_0 = 0$ and $p_0 = 0$. Under this logit specification, firm i 's market share is equal to (McFadden (1974))

$$s_i(\omega, p) = \frac{\exp(g(\omega_i) - p_i)}{1 + \sum_j \exp(g(\omega_j) - p_j)}.$$

Firms derive flow profits from static competition in prices for the market shares above. There is a unique Bertrand-Nash equilibrium characterized by the FOCs (Caplin and Nalebuff

³⁵For example, see Pakes and McGuire (1994), Pakes and McGuire (2001), and Doraszelski and Satterthwaite (2010) among others for advances in equilibrium computation for the discrete-time version of the model.

(1991))

$$-(p - mc)(1 - s) + 1 = 0,$$

where mc is a constant marginal cost. Solving these FOCs gives prices p as a function of the states ω . Note the resulting profit function $\pi(\omega_1, \dots, \omega_i, \dots, \omega_N; \theta_\pi)$ is increasing in ω_i and decreasing in the other elements.

The firm's only dynamic control is on investment $\kappa_i > 0$, which is subject to cost $c(\kappa_i; \theta_c) = \frac{1}{2}\theta_c\kappa_i^2$. Investment counteracts the natural decay on quality (representing the increasing appeal of an outside good), so the law of motion for quality is

$$d\omega_i = \mu_{\omega_i} dt = (-\delta + \kappa_i) dt,$$

where δ is the decay rate. The resulting HJB is then

$$\rho V_i(\omega) = \pi_i(\omega; \theta_\pi) + \max_{\kappa_i} (\mu_{\omega_i}(\kappa_i) V_{i,\omega_i}(\omega) - c(\kappa_i; \theta_c)) + \sum_{j \neq i} \mu_{\omega_j} V_{i,\omega_j}(\omega).$$

The firm's optimal choice of investment is characterized by the FOC

$$V_{i,\omega_i}(\omega) = c'(\kappa_i; \theta_c),$$

which implies $\kappa_i^*(\omega) = V_{i,\omega_i}(\omega) / \theta_c$.

2.5.2 Numerical Results

As mentioned earlier, we focus on the case with $N = 2$ to facilitate computation of a dense benchmark solution. We assume that $\rho = 0.05$, and set the other parameters to $\delta = 0.25$, $\delta_\pi = 10$, $\theta_c = 15$, $mc = 1$. We start by solving the model on a large dense grid of 66,049 nodes and use this solution as a benchmark for evaluating the error rates of dense and adaptive sparse grids (adapting iteratively on the solution for the value function) of varying sizes. While our toolbox is designed for much higher N , we purposefully consider the duopoly case because it is simple enough to obtain a benchmark solution via a dense grid. Since sparsity is most impactful in higher dimensionality, the performance we show here likely understates computational gains under general use cases.

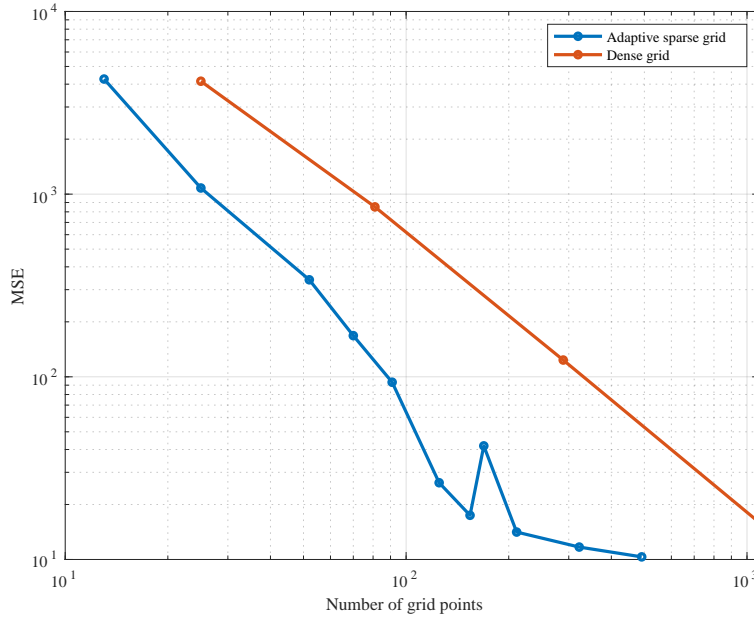


Figure 2.9: Error rates for quality ladder model

Figure 2.9 plots the error rates for the solution for the value function. For each grid, we compute the implied value function over all 66,049 nodes of our benchmark grid and obtain the MSE over these 66,049 nodes. Across grids of varying size, adaptive grids outperform dense grids by roughly one order of magnitude.

2.6 Conclusion

In this paper, we develop a scalable and portable method to solve high-dimensional dynamic programming problems in continuous time using adaptive sparse grids. Our main use cases are heterogeneous-agent models from different fields in economics, which can be represented in continuous time as systems of partial differential equations. We showcase the power of our method by solving high-dimensional variants of the seminal Krusell and Smith (1998) and Ericson and Pakes (1995) models in continuous time. Our paper is accompanied by a numerical toolbox with automated general-purpose routines to solve dynamic programming problems in continuous time, thus hopefully allowing for broad accessibility.

Chapter 3

A Theory of Dynamic Inflation Targets¹

3.1 Introduction

The policy discussion has increased attention on the question of whether central bank inflation targets should be perpetually fixed, as in the US, or whether they should be adjusted. For example, a central bank may wish to adjust its target upward during normal times to bring itself away from the zero lower bound and allow for a larger monetary policy response during a recession.² Given that inflation targets are designed to anchor inflation expectations, this raises the natural concern that allowing a flexible target would undermine its purpose. Moreover, there are different methods by which adjustment could occur, with possibilities including allowing adjustment at fixed points in time or for flexible adjustment within a restricted band.

This paper provides a framework to study whether inflation targets should be adjusted over time, and if so what the process governing target adjustment should be. To do so, we study a mechanism design problem of a government designing a transferable utility mechanism to control how a central bank sets monetary policy (inflation). The central bank has private information over persistent structural economic shocks which determine the optimal rate of inflation, meaning that optimal inflation can vary systematically over time. However, the central bank also has a time consistency problem in inflation setting that arises from a Phillips Curve

¹Co-authored with Christopher Clayton

²See e.g. Ball *et al.* (2016).

relationship. This results in a trade-off between flexibility to allow monetary policy to respond to structural shocks, and commitment to manage the time consistency problem. Additionally, information persistence implies that when the central bank reveals its information, it affects the beliefs of both the government and of the firms that determine the inflation-output relationship.

The main result of this paper is that a dynamic inflation target implements the constrained efficient level of inflation, and moreover is the optimal mechanism when the social cost of the transfer (enforcement) mechanism governing central bank behavior is negligible relative to the inflation-output trade-off. Under this mechanism, the central bank's inflation setting in each period is managed by the target, with linear incentives for deviations of inflation from the target. Moreover, central bank can update the target *one period in advance*. As a result, in every period the central bank both sets inflation based on the existing target, and also decides whether to update the target for the next period. In equilibrium under this mechanism, the target is always equal to next-period expected inflation, giving rise to the interpretation as a dynamic inflation target.

The intuition behind the optimality of the dynamic inflation target is as follows. First, the target itself manages the time consistency problem in inflation setting by penalizing inflation in excess of the target. Second, requiring the target to be adjusted one period in advance manages the time consistency problem in target adjustment. If the central bank were allowed to *contemporaneously* adjust its inflation target, it would adjust its target to its desired inflation level, de-anchoring inflation expectations and rendering the target meaningless. By adjusting in advance, the central bank internalizes its future time consistency problem and sets the target efficiently. This allows inflation expectations to remain anchored to the target, even though it is adjusted over time. Moreover, the information effects of changing the target on the beliefs of the government and firms offset each other at the constrained efficient allocation due to the fact that both firm behavior and government transfers are based on inflation expectations.

The optimality of a dynamic inflation target suggests that controlled target adjustment may be preferable to a perpetually fixed target, and moreover provides a means of controlling target adjustment. It suggests that adopting an approach similar to that of the Bank of Canada, where the target can be adjusted only at fixed five year intervals, may be an effective way of allowing

for target adjustment without de-anchoring inflation expectations.

We explore the robustness of our results to costly enforcement (transfers) - that is, the mechanism used to control central bank behavior has first order welfare costs. Although the dynamic inflation target still implements the constrained efficient allocation, the government no longer finds it exactly optimal to do so due to the enforcement costs. Nevertheless, we show that the properties of the optimal mechanism still bear features resembling a dynamic inflation target.

Related Literature: We related to the literature on time consistency in monetary policy, as well as to the broader mechanism design literature on the trade-off between commitment and flexibility.³ Walsh (1995) shows that an inflation target is an optimal mechanism in a static context with transferable utility. Halac and Yared (2019) considers the trade-off between instrument-based rules and target-based rules in a delegation framework. Several papers study the commitment-flexibility trade-off in a delegation framework (Amador *et al.* (2006), Athey *et al.* (2005)), including with persistent information (Halac and Yared (2014)). Beshears *et al.* (2020) considers the trade-off in a population with present bias and inter-agent transfers. Our contribution is to study the problem of inflation setting in a transferable utility framework with both persistent information and outside agents who derive information from reports under the mechanism, and to show its implications for inflation target adjustment. In our model, there are two commitment-flexibility trade-offs: first in allowing for present inflation to respond to shocks (“the target”), and second in allowing for adjustment of future targets to shocks (“changing the target”).

3.2 Model

Our economy has a government, a monetary authority (central bank), and a continuum of small firms. Time is infinite and discrete, indexed by $t = 0, 1, \dots$. Allocations in this economy are summarized by two scalar variables, inflation π_t and output y_t . We introduce these labels

³See e.g. Barro and Gordon (1983), Canzoneri (1985), Kydland and Prescott (1977), Persson and Tabellini (1993), and Rogoff (1985) for the former.

from the start to allude to the standard New Keynesian model, but everything that follows extends to the general case where π_t and y_t are vectors. Both inflation and output lie in compact sets, $\pi_t \in [\underline{\pi}, \bar{\pi}] \subset \mathbb{R}$ and $y_t \in [\underline{y}, \bar{y}] \subset \mathbb{R}$. The relationship between inflation and output is determined by price-setting firms.

The government and central bank interact in a principal-agent framework. The central bank learns about the state of the economy, and uses this information to determine monetary policy. However, the central bank is subject to a time-consistency problem. As a result, the government will design a mechanism to determine how the bank sets monetary policy, subject to truthful reporting conditions. Firms are not directly under the control of the government, so their actions will be taken as a constraint on the problem.

Government. The social preferences of the government are given by

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \mathcal{U}_t(\pi_t, y_t, \theta_t), \quad (3.1)$$

where β is the discount factor, and $\theta_t \in \Theta = [\underline{\theta}, \bar{\theta}]$ denotes an economic shock with conditional density $f(\theta_t | \theta_{t-1})$. The government does not directly observe the shock θ_t , which will be observed and reported by the central bank.

Interpretation of θ_t . We think of θ_t as corresponding to the true economic shock to the economy, which the central bank learns information about. In this sense, the government welfare function reflects a notion of social welfare that incorporates the true shock - for example, a true supply side shock or shock to the loss function for inflation and output deviations. However, both the firms and the government rely on the central bank to collect and report information about that true type. As such, both the government and firms form decision rules based on the *reported* type, not the true type. The government's beliefs matter because it designs the mechanism that maps the reported type (beliefs) into an allocation rule. The firm's beliefs also matter, because firms form a decision rule based on those beliefs. In both cases, that decision rule responds to beliefs, while the "true" welfare function is related to the true type.

Central Bank. The central bank has preferences over both social welfare, and over transfers T_t from the government, so that its preferences are

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[\mathcal{U}_t(\pi_t, y_t, \theta_t) + T_t \right]. \quad (3.2)$$

The central bank observes the shock realization θ_t in every period, and so is tasked with choosing inflation and output in each period.

In writing preferences, we have assumed that T_t is welfare neutral from the perspective of the government, and is only used as a control mechanism. This can be seen as a limiting case where the costs of controlling central bank behavior are negligible relative to the underlying social welfare problem. In Sections 3.4 and 3.5, we study the case where transfers are not neutral from the perspective of the government.

Interpretation of Transfers. Rather than monetary transfers, the practical analogs of the control mechanism T_t may be closer to policies such as Congressional scrutiny, reputational risk, or firing the central banker.⁴ For example, a central bank being awarded high T_t may face a low degree of Congressional scrutiny in its policy determination. So far, T_t could be a utility transfer, or it could be a form of money burning.

Firms and Phillips Curve. Firms, whose actions are not directly controlled by the government or central bank, create a “Phillips Curve” relationship between inflation and output, given by

$$y_t = F_t(\pi_t, \mathbb{E}_t[\pi_{t+1} | \mu_t]). \quad (3.3)$$

The object μ_t is the belief of firms about the current state θ_t , which are formed from the report of the central bank. Because θ_t will feature persistence, beliefs about the current state of θ_t translate into beliefs about the next period state θ_{t+1} and hence the next-period inflation π_{t+1} . Because firms cannot be directly controlled, equation (3.3) is an implementability condition from the

⁴In the U.S., for example, this process is multifaceted. The central bank Chairwoman is directly held accountable to Congress in the form of bi-annual, as well as extraordinary, Congressional testimonies. The institutions adopted by modern central banks to allow for active monitoring by stakeholders and maintain accountability vary across countries but most are highly complex.

perspective of the government and central bank.

To simplify exposition, we internalize the Phillips Curve relationship of equation (3.3) into preferences, and write

$$u_t(\pi_t, \mathbb{E}_t[\pi_{t+1}|\mu_t], \theta_t) = U_t(\pi_t, \pi_{t+1}, \theta_t, \mu_t) = \mathcal{U}_t(\pi_t, F_t(\pi_t, \mathbb{E}_t[\pi_{t+1}|\mu_t]), \theta_t)$$

where we use U_t and u_t for more compact notation.

Lucas Critique. A key concern of this Phillips Curve relationship is a Lucas critique - firm price setting behavior may change in response to inflation policy changes, such as target changes.⁵ Our Phillips Curve relationship is robust to a Lucas critique provided that expected future (next period) inflation is sufficient for determining how changes in future policies affect firm behavior. For example, higher expected inflation may lead firms to increase the frequency with which they update prices, altering the slope of the Phillips Curve.

3.2.1 Constrained Efficient Allocation

Before characterizing the mechanism structure, we characterize the constrained efficient allocation that could be achieved if the government and firms observed θ_t , and the government mandated with commitment the inflation decisions of the central bank. This provides an efficiency benchmark which respects the Phillips Curve relationship between inflation and output determined by firms.

The constrained efficient planning problem of the government is given by the optimization problem

$$\max_{\{\pi_t\}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}_0 \left[u_t(\pi_t, \mathbb{E}_t(\pi_{t+1} | \theta_t), \theta_t) \right], \quad (3.4)$$

where $\pi_t = \pi_t(\theta^t)$ is adapted to the full history of realized shocks, and where θ_{-1} is an initial condition of the model. Taking the first order conditions, we obtain the necessary conditions for

⁵See e.g. L'Huillier and Schoenle (2019).

optimality,

$$\frac{\partial u_t}{\partial \pi_t} = -\frac{1}{\beta} \frac{\partial u_{t-1}}{\partial \mathbb{E}_{t-1}(\pi_t | \theta_{t-1})} \quad \forall t \geq 1 \quad (3.5)$$

$$0 = \frac{\partial u_0}{\partial \pi_0} \quad (3.6)$$

Because there is no Phillips curve constraint for π_0 (it would occur at date -1), the optimality condition for π_0 simply sets the marginal value of increasing inflation to 0, yielding a standard first-order condition that would also be chosen by the central bank. For $t \geq 1$, the first-order condition reflects the Phillips curve relationship, and internalizes that inflation at date t affects output at date $t - 1$ through the constraint set. The LHS of equation (3.5) is date t -adapted, whereas the RHS is date $t - 1$ -adapted. Therefore, the RHS is constant from the perspective of time t , implying that the marginal (flow) utility from inflation is constant at date t in histories θ^t proceeding from the same history θ^{t-1} .

Discretionary monetary policy. Suppose that the central bank were left to its own devices to set inflation ($T_t = 0$) with discretion. At period t for any t , the central bank finds it optimal to set $\partial u_t / \partial \pi_t = 0$ state by state. In particular at date t , the central bank neglects the impact of inflation on the previous period's Phillips Curve, which no longer serves as a constraint of the problem. This results in inflationary bias and reflects a standard Barro and Gordon (1983) time consistency problem. This motivates studying a mechanism to control central bank inflation setting.

A Sufficient Statistic. Suppose that we evaluate the constrained efficient allocation, and define the wedge

$$v_{t-1} = -\frac{1}{\beta} \frac{\partial u_{t-1}}{\partial \mathbb{E}_{t-1}(\pi_t | \theta_{t-1})}. \quad (3.7)$$

This wedge v_{t-1} is a date $t - 1$ adapted constant that is a sufficient statistic for the shock history θ^{t-1} in determining the allocation rule π_t, π_{t+1}, \dots for inflation. v_{t-1} is the wedge between the constrained efficient allocation rule for π_t and the central bank's allocation rule under discretion. In other words, this wedge reflects the inflationary bias that arises from the central bank's time consistency problem.

3.2.2 Mechanism Structure

We study direct and *full-transparency* mechanisms, under which the central bank makes a report of its type each period.⁶ By full transparency, we mean that there is no pooling of central bank types in reporting. We denote this report by $\tilde{\theta}_t$. Along the equilibrium path, agents' posterior will therefore be the degenerate distribution at the reported type, or $\mu_t = \tilde{\theta}_t$. Note that we abuse notation here because μ_t is a full distribution in general.

Restricting attention to *full transparency* mechanisms is not without loss of generality. In principle, the government could want to pool central bank types to shroud the type and manipulate firm beliefs. In assuming mechanisms under which the central bank truthfully reveals its type, we assume away such motivations. Given that central bank transparency has received increasing emphasis since the financial crisis, we impose the restriction to full transparency.⁷

A mechanism in our model is a mapping from the history of *reported* types into a transfer and allocation rule, given by $(\pi_t, T_t) : \Theta^t \rightarrow \mathbb{R}^2$. Although the date t allocations depend on the entire history of reported types, we will characterize state space reduction results which will allow us to characterize sufficient statistics for information histories.

3.2.3 Incentive Compatibility, Time Consistency, and Information

At every date t , the central bank makes a report $\tilde{\theta}_t$ of its true type θ_t , subject to incentive compatibility. We characterize the local incentive compatibility constraints used in solving the relaxed problem. We will apply a *first order approach* to solving the problem.⁸ We use incentive compatibility to identify three driving forces of the model: a time consistency problem from the Phillips Curve, and two informational problems related to shock persistence.

Suppose that we define the value function of the central bank by $\mathcal{W}_t(\theta^t)$, where θ^{t-1} is the

⁶Note that under an indirect mechanism, e.g. a report $\sigma(\theta)$ for some rule σ , the true type can be uncovered by unwinding the indirect report in the usual manner.

⁷See e.g. Powell (2019).

⁸We make use of techniques developed in Pavan *et al.* (2014) for a first order approach with persistent shocks. We do not provide or verify a monotonicity condition that can be used to verify that the first order approach yields a globally incentive compatible mechanism.

history of *reported* types whereas θ_t is always the current true type. The central bank's reporting problem at date t is given by

$$\mathcal{W}_t(\theta^t) = \max_{\tilde{\theta}_t} \mathbb{E}_t \left[\sum_{s=t}^{\infty} \beta^s (U_s(\pi_s, \pi_{s+1}, \theta_s, \tilde{\theta}_s) + T_s) \mid \theta_t \right]$$

where the history-dependence is implicit in the policies (π_s, T_s) . The current report $\tilde{\theta}_t$ affects not only the current allocations (π_t, T_t) , but through history dependence also affects all future allocations. In addition, it also affects firm beliefs $\tilde{\theta}_t$. Formally, we can write the local incentive compatibility condition as

$$\mathbb{E}_t \sum_{s=t}^{\infty} \left[\beta^{s-t} \frac{\partial T_s}{\partial \tilde{\theta}_t} \mid \theta_t \right] = -\mathbb{E}_t \left[\sum_{s=t}^{\infty} \beta^{s-t} \left(\frac{\partial U_s}{\partial \pi_s} \frac{\partial \pi_s}{\partial \tilde{\theta}_t} + \frac{\partial U_s}{\partial \pi_{s+1}} \frac{\partial \pi_{s+1}}{\partial \tilde{\theta}_t} \right) + \underbrace{\frac{\partial U_t}{\partial \tilde{\theta}_t}}_{\text{Firm Beliefs}} \mid \underbrace{\theta_t}_{\text{Persistent Shocks}} \right]$$

Local incentive compatibility gives the change in total transfers required to implement an allocation. The change in transfers depends on changes in the allocation rule, which reflect both the contemporaneous utility impact and the impact through the Phillips Curve relationship.

There are three key forces that appear in this decision rule: one reflecting time consistency, and two reflecting information considerations related to persistent shocks.

The first, reflecting time consistency, is the *absence* of the impact on the previous period's Phillips Curve. This reflects the Barro and Gordon (1983) time consistency logic, that the central bank neglects the previous period Phillips Curve when determining current inflation. This is the first force that the government must account for when designing the mechanism.

The second force, related to persistent shocks, is that the central bank evaluates expectations with respect to the *true* type θ_t , whereas the government designs policy and forms expectations based on the *reported* types. This creates an informational friction where the central bank has an incentive to manipulate the beliefs of the government, while its own beliefs are held constant.

The third and related force is that an analogous information manipulation incentive exists for firm beliefs. Firms form beliefs based on the reported type, meaning that the central bank can manipulate their actions with the information it reports. This appears through the appearance of inflation expectations of firms in the Phillips Curve relationship, which are formed based on their beliefs. In the conventional New Keynesian framework, the central bank will want to bias

firm inflation expectations *downward* in order to improve the contemporaneous inflation-output trade-off.

These three forces affect not only the central bank's inflation decision, but also its reporting decision. When designing a mechanism, the government must account for all three of these effects.

3.2.4 Naive Target Adjustment

It is well known that an inflation target is a method of controlling inflationary bias. This proposal arises in static mechanism design problems as a way to enforce efficient inflation policies. Relative to the static setting, however, we feature a dynamic problem of persistent information, where the optimal inflation rate moves substantially over time, and where there are additional informational incentives over the beliefs of firms and the government. As a result, a static target would not be able to achieve the constrained efficient level of inflation, and is unlikely to constitute an optimal mechanism.

However, motivated by the previous literature, a benchmark proposal would be to simply grant the central bank flexibility to adjust its target. This would allow the central bank to accommodate structural shocks that change the inflation-output trade-off in a fundamental way. It is clear, however, that this type of mechanism would simply reintroduce the Barro and Gordon (1983) time consistency problem and undermine the target. The central bank would have an incentive in every period to evaluate its optimal policy $\frac{\partial u_t}{\partial \pi_t} = 0$, and then to set its inflation target to coincide with that inflation level. Rather than only resetting its target in response to structural shocks, the central bank also resets its target to accommodate its inflationary bias.

However, recognizing that the central bank would set its inflation target each period to coincide with its discretionary policy, firms in the previous period would form inflation expectations that coincided with expected inflation under discretion. In other words, firm inflation expectations are no longer anchored to the current target because they know it will simply be adjusted in the next period to accommodate the inflationary bias. This results in a de-anchoring of inflation expectations. Naive adjustment undermines the target by allowing the central bank to set its target to its desired inflation, rather than the intended goal of forcing

inflation to match the target.

3.3 Dynamic Inflation Target

We show that a “dynamic inflation target” mechanism can implement the constrained efficient allocation when the target is set *one period in advance*.

In particular, equations (3.5) and (3.7), along with their sufficient statistic implications, suggest the following mechanism: use the transfer rule T_t to penalize deviations from a target. In other words, we look over a class of mechanisms defined by the rule

$$-b_{t-1}(\pi_t - \tau_{t-1}), \quad (3.8)$$

where τ_{t-1} is the target, and b_{t-1} is the slope of the punishment for inflation in excess of the target. A target of this form seeks to correct the time consistency problem in central bank inflation setting by incentivizing the central bank to set inflation to its target.

However, in the presence of structural shocks, the target might need to be adjusted over time to accommodate a changing efficient level of inflation. A persistent structural shock may drift the optimal inflation rate far from the present target in a persistent manner, resulting in larger gains from letting the target adjust and suggesting a structural change in the commitment-flexibility trade-off.

As a result, in addition to choosing inflation, the central bank must decide each period whether and how to adjust the target. Under our proposed mechanism, the central bank adjusts its target for the *following* period. It does so by reporting its type $\tilde{\theta}_t = \theta_t$, which generates a mapping $\tilde{\theta}^t \rightarrow (b_t, \tau_t)$ from the history of reported types into the next period target. In practice, we can also think of the central bank as directly choosing its own target, represented by (b, τ) , from among a set of pairs.

In sum, we propose the following mechanism. In each period, the central bank chooses inflation π_t and reports its type to determine the next period target. We show that this “dynamic inflation target” implements the constrained efficient level of inflation in a locally incentive compatible mechanism, which moreover admits a key state space reduction property.

Proposition 4 (Dynamic Inflation Targeting). *A dynamic inflation target implements the constrained efficient allocation in a locally incentive compatible mechanism. The mechanism is represented by the Bellman equation*

$$\mathcal{W}(v_{-1}, \mathbb{E}_{-1}(\pi), \theta) = \sup_{\pi, \tilde{\theta}} \left\{ \underbrace{-v_{-1}[\pi - \mathbb{E}_{-1}(\pi)] + U(\pi, \mathbb{E}(\pi_{+1} | v_{-1}, \tilde{\theta}), \theta)}_{\text{Dynamic inflation target}} \right. \quad (3.9)$$

$$\left. + \beta \mathbb{E} \left[\mathcal{W}(v(v_{-1}, \tilde{\theta}), \mathbb{E}(\pi_{+1} | v_{-1}, \tilde{\theta}), \theta_{+1}) \mid \theta \right] \right\},$$

where v and $E(\pi_{+1})$ are given by their constrained efficient values. v_{-1} is a sufficient statistic for the history $\tilde{\theta}^{-1}$ of reported types.

Proposition 4 shows that in a setting with persistent structural shocks, the optimal mechanism collapses to a simple dynamic inflation target where inflation always meets the target in expectation, that is $\tau_{-1} = \mathbb{E}_{-1}\pi$. The slope of the target mechanism is always given by $b_{-1} = v_{-1}$.

Our mechanism features a time-varying inflation target, where each period the time consistency in inflation setting is corrected by the target. As in the static model, the marginal welfare cost of future inflation is always through expectations in the Phillips curve, and so is constant on the margin across all future states. As a result, a linear inflation penalty for exceeding the target corrects the time consistency problem. This component of the mechanism is analogous to standard inflation targeting logic (e.g. Walsh (1995)) from the static setting.

However, the inflation target also needs to be adjusted in response to structural shocks in order to allow inflation to coincide with its constrained efficient rule. The key insight of Proposition 4 is that the central bank optimally resets this target one period in advance. In essence, if the central bank observes a persistent shift in the constrained efficient inflation rate from 2% to 3%, the correct response of the central bank is to adjust the inflation target for the *next* period from 2% to 3%.

Moreover, the current target $(v_{-1}, \mathbb{E}_{-1}\pi)$ is a sufficient statistic for the history $\tilde{\theta}^{-1}$ of reported types. This sufficient statistic property follows precisely as in the constrained efficient allocation, as the slope of the target v_{-1} reflects the previous period time consistency problem.

This implies that all history dependence of the mechanism is captured in the inflation target, and greatly reduces the knowledge required to adjust the target.

Target adjustment is subject to all three frictions identified above: time consistency, government belief manipulation, and firm belief manipulation. A naive inflation target adjusted contemporaneously resulted in an un-anchoring of inflation expectations, as the central bank could simply set its target to its desired inflation. By requiring instead the target to be set one period in advance, the central bank overcomes this time consistency problem. When setting the target at date $t - 1$, the central bank accounts for the current Phillips Curve relationship, overcoming the time consistency problem.⁹

If shocks were not persistent, time consistency would be the only force in the model, and Proposition 4 would still characterize optimal policy. This reflects the logic of the static model, with the additional insight that by setting the target in advance, the time consistency problem in target adjustment is overcome. In this sense, the optimality of the dynamic inflation target directly reflects the logic of the standard time consistency problem. However, if there is no shock persistence and shocks are small, there may be little value to adjustment the target over time, that is little commitment versus flexibility trade-off.

By contrast when shocks are large and persistent, there may be a strong incentive to adjust the target. At the same time, however, two additional incentive problems emerge: government and firm belief manipulation. It is therefore surprising that this simple target and adjustment mechanism remains relevant. In the constrained efficient allocation, these two informational effects exactly offset one another. In particular, when the central bank considers on the margin changing its report $\tilde{\theta}_t$, this generates a welfare impact¹⁰

$$\underbrace{\frac{\partial u}{\partial E(\pi_{+1} | \tilde{\theta})} \frac{dE[\pi_{+1} | \tilde{\theta}]}{d\tilde{\theta}}}_{\text{Firm Beliefs}} + \underbrace{\beta v \frac{dE[\pi_{+1} | \tilde{\theta}]}{d\tilde{\theta}}}_{\text{Government Beliefs}}$$

Under the constrained efficient allocation, we have $\frac{\partial u}{\partial E(\pi_{+1} | \tilde{\theta})} + \beta v = 0$. In economic terms, on the

⁹This is similar to the static setting, where the central bank is willing “ex ante” to set up a targeting mechanism for itself. It is also closely related to the literature on optimal mechanisms to control present bias (e.g. Amador *et al.* (2006)), where agents are willing to set up mechanisms to control their own time consistency problems.

¹⁰See the proof of Proposition 4 for a detailed derivation.

one hand the central bank wishes to bias inflation expectations of the government *upward*, which increases the target and reduces penalties for exceeding its target. On the other hand, the central bank also wishes to bias the inflation expectations of firms *downward*, in order to economize on the Phillips Curve relationship and improve the contemporaneous inflation-output trade-off. At the constrained efficient allocation where the slope of the target is exactly equal to the Phillips Curve impact, the marginal reduction in penalties for exceeding the target is exactly equal to the Phillips Curve impact, and these two forces exactly offset one another.

It is worth remarking two forces exactly cancel each other because central bank and government align at date $t - 1$ over the future inflation-output trade-off. In fact, cancellation occurs by construction. The central bank at date $t - 1$ agrees with the government about optimal inflation from dates t and onward, internalizing the Phillips Curve relationship, but also wishes to manipulate firm beliefs. As a result, the government designs its transfer rule to offset firm belief manipulation. This results in these two informational effects exactly canceling one another. In a more general setup with preference disagreement, the government would also have to account for differences in preferences in controlling the reported type, leading to additional forces in the transfer rule. Moreover, when considering the optimal mechanism when transfers are costly (Section 3.5), counteracting firm belief manipulation will be costly to the government, leading to distortions in the allocation rule relative to constrained efficiency.

Dynamic Inflation Target as an Optimal Mechanism. Proposition 4 makes a statement about implementability: a dynamic inflation target *can* implement the constrained efficient allocation of inflation. It does not state that it is an optimal mechanism. It is immediate that if we assume the control mechanism has no direct welfare consequence to the government – that is, T_t does not appear in the government objective function – then the dynamic inflation target is also an optimal mechanism, because it maximizes the government’s objective function. This implies that if the government’s primary concern is with the inflation-output trade-off, and not with the costs of implementing the mechanism, the dynamic inflation target is optimal.

3.3.1 Evolution of the Target

The characterization of the target in Proposition 4 can be used to understand the evolution of the target. The first order condition for optimal inflation choice implies that $v_{t-1} \frac{\partial U_t}{\partial \pi_t} + \frac{\partial U_t}{\partial y_t} \frac{\partial F_t}{\partial \pi_t}$. Combining this with the definition of the updated target, $v_t = -\frac{1}{\beta} \frac{\partial U_t}{\partial y_t} \frac{\partial F_t}{\partial E_t \pi_{t+1}}$, and making the natural assumption that $\frac{\partial F_t / \partial \pi_t}{\partial F_t / \partial E_t \pi_{t+1}} < 0$ local to the constrained efficient allocation,¹¹ we obtain a law of motion,

$$\Delta v_t = \underbrace{-\frac{\partial U_t}{\partial \pi_t}}_{\text{Excess Inflation}} + \underbrace{\left(1 - \left| \frac{\partial F_t / \partial \pi_t}{\partial F_t / \partial E_t \pi_{t+1}} \beta \right| \right)}_{\text{Time Consistency}} v_{t-1}, \quad (3.10)$$

reflecting the extent to which the slope of the target updates.

The first effect, “Excess Inflation,” indicates that the slope of the target *increases* as the central bank reduces marginal utility from inflation. In conventional models, this corresponds to higher levels of inflation. In absence of a persistent structural shock, this term reflects a conventional New Keynesian Phillips Curve (NKPC) logic. When the central bank generates excess inflation in order to stimulate output, the central bank also has a motivation to *reduce* future inflation in order to *reduce* future inflation in order to capitalize on the current inflation-output trade-off. This leads the central bank to increase the slope of the target, punishing future inflation exceeding the target and reducing inflation expectations.

The second force, “Time Consistency,” reflects the relative effects of current and future inflation on current output through the Phillips Curve. In the standard NKPC, $\left| \frac{\partial F_t / \partial \pi_t}{\partial F_t / \partial E_t \pi_{t+1}} \beta \right| = 1$, and this term drops out. More generally, when the relative effect of current inflation is large relative to future inflation, that is $\left| \frac{\partial F_t / \partial \pi_t}{\partial F_t / \partial E_t \pi_{t+1}} \beta \right| > 1$, the time consistency problem is small relative to demand stabilization, leading to a reduction in the slope of the target. By contrast when $\left| \frac{\partial F_t / \partial \pi_t}{\partial F_t / \partial E_t \pi_{t+1}} \beta \right| < 1$, the time consistency problem is relatively severe, and the slope of the target is increased.

Suppose then that a persistent shock structurally increases the optimal rate of inflation. As a result, the slope of the target changes following a structural shock only if either the incentive to generate *excess* inflation changes, or the slope of the relative slopes of the Phillips Curve

¹¹This assumption implies that current and future inflation move current output in opposite directions.

fundamentally change. This implies that if a persistent structural shock increase the optimal rate of inflation, the slope of the target increases (decreases) only if it either increases (decreases) the motivation to generate *excess* inflation, or if it exacerbates (dampens) the time consistency problem from the Phillips Curve. Even though a persistent shock may lead to an increase in the target (expected inflation), it may or may not lead to a change in the slope of the target.

Finally, we can consider how the target intercept changes in response to a structural shock. Considering a marginal increase in the structural shock, we have

$$\frac{d\tau_{t-1}}{d\theta_{t-1}} = \underbrace{E_{t-1} \left[\pi_t \frac{\partial f(\theta_t|\theta_{t-1})/\partial\theta_{t-1}}{f(\theta_t|\theta_{t-1})} \right]}_{\text{Expectations}} + \underbrace{\frac{\partial v_{t-1}}{\partial\theta_{t-1}} E_{t-1} \left[\frac{\partial\pi_t}{\partial v_{t-1}} \middle| \theta_{t-1} \right]}_{\text{Target Slope Adjustment}}. \quad (3.11)$$

There are two effects of the structural shock on the target. The first effect, “Expectations,” reflects that the probability measure over future states changes in response to the structural shock. If an increase in the state θ_t increases the probability of high-inflation states, then it increases the target intercept, τ_{t-1} . The expectations effect implies that the target intercept changes even in absence of a change in the target slope in the presence of a structural shock. Notice that if the shock were transitory and did not affect the probability measure, it would not lead to an adjustment of the target intercept. The second effect, “Target Slope Adjustment,” reflects the extent to which a change in the slope of the target impacts optimal future inflation. In the natural case where $\frac{\partial\pi_t}{\partial v_{t-1}} < 0$, the upward adjustment of the target intercept is *amplified (dampened)* when the target slope is reduced (increased).

In sum, when the economy experiences a structural shock θ_t , both components of the target may be affected. The intercept of the target is directly affected by changes in expectations, but is also affected indirectly if the structural shock leads to a change in target slope. The target slope updates as a result of the structural shock if it leads to a fundamental change in either the motivation of the central bank to generate excess inflation, or if it alters the nature of the time consistency problem.

3.3.2 On the Nature of Inflation Targets

The dynamic inflation target of Proposition 4 sheds light on the important features of the target in anchoring inflation expectations. In particular, the target's efficacy does not result from it being constant and unchanging, but rather from it being committed to sufficiently in advance. De-anchoring of expectations arises in the naive adjustment process because the central bank is able to match its target to its desired inflation, rather than forcing inflation to match the target. By ensuring target setting in advance, the central bank is forced to set its inflation to its target instead.

3.3.3 Target Adjustment in Practice

Proposition 4 provides a mechanism to implement the constrained efficient allocation, which takes the form of a dynamic inflation target. The key property of the target adjustment process is that it occurs one period in advance.

The inflation target adjustment in our model happens at the frequency of the underlying structural shock, in the sense that large movements in the target will generally coincide with large and persistent shock realizations. This suggests that, in practice, target adjustment would either require the new target to either be announced sufficiently far in advance of it taking effect, or the target being adjusted contemporaneously but for a long enough interval of time. This latter method minimizes the time consistency problem by forcing the target to remain in force for a duration of time, so that the central bank is guided primarily by the long-term consequences of the target and not by short-term inflationary bias.

What does a *period in advance* mean in practice? In the model, the key property of a period is the expectations appearing through the Phillips Curve, which generates the time consistency problem. This relationship arises traditionally because firms adjust prices only infrequently. One notion of a period in advance, then, would be a notion of the duration of firm prices resetting. The target adjustment should happen sufficiently far in the future that a sufficient number of firms will have had the chance to reset prices, in order to minimize time inconsistency. In other words, as the frequency of price adjustment increases, the interval of time constituting a "period"

falls.

A related force is the lag between central bank policy setting and its impact on output. As this lag increases, the effective frequency of price resetting also increases from the perspective of time consistency. A longer lag would reduce the interval of time constituting a period.

On the other hand, it is also important to know the duration of structural shifts in the underlying shock. When the frequency of structural shocks is low relative to the frequency of price adjustment, a central bank observing a structural shock does not expect to observe another one until well after firm prices have adjusted. We would therefore expect lower frequency structural shocks to increase the notion of period length.

The notion of the “correct” length of a period must consider all three of these forces. Fundamentally, the trade-off reflects a form of commitment versus flexibility trade-off in target adjustment. On the one hand, granting greater flexibility (more frequent target adjustments) allows responses to structural shocks. On the other hand, greater flexibility also exacerbates the time consistency problem.

Comparing our mechanism to real-world inflation target regimes. A close analogue to our mechanism in practice is the adjustment process for the Bank of Canada.¹² The Bank of Canada revisits its inflation target every five years at fixed intervals, with the possibility of adjusting its target for the next five-year interval based on new information. Our results suggest that this type of adjustment process can implement the efficient level of inflation by providing a means of responding to structural economic shocks without generating substantial time consistency problems in target setting.

3.3.4 Relation to Conservative Central Banker

Although inflation targets are one solution to time consistency problems, appointment of a conservative central banker is another possible solution.¹³ The linear penalty for inflation around expectations could be understood as a time-varying central banker, where rather than increasing

¹²For background, see <https://www.bankofcanada.ca/core-functions/monetary-policy/inflation/>.

¹³See Rogoff (1985).

(decreasing) the target, control of the central bank is handed over to a more dovish (hawkish) central banker. Under this interpretation, higher ν_{-1} reflects more hawkish central bankers.

This interpretation requires a clearer interpretation of $\nu\mathbb{E}\pi$. Under this interpretation, $\nu\mathbb{E}\pi$ is constant except when transitioning between central bankers types θ_t , when it jumps. This suggests that implementing the mechanism in this form requires some notion of scrutiny of a central bank when it decides to transition power.

3.4 Inflation Targeting and Reputation

Section 3.3 showed that a dynamic inflation target implements the constrained efficient allocation, and therefore is optimal when the cost of the control mechanism is negligible.

In addition to the interpretation as a conservative central banker offered in Section 3.3, there are two additional interpretations consistent with a negligible social cost.

The most direct interpretation is as a monetary compensation scheme. In this interpretation, T corresponds to direct “incentive” payments to the central banker for inflation outcomes. Such payments would be small as a fraction of the economy, suggesting that zero social cost of the scheme is a plausible approximation. Moreover, we show in Section 3.5 that even when there are social costs to these payments, the optimal mechanism has properties resembling the dynamic inflation target of Section 3.3.

A second interpretation of T is that it corresponds to the “reputation” of the central bank or central banker. Under this interpretation, the central bank announces not only its target but also its “commitment” to the target, the latter of which is represented by the slope. A central bank which exceeds its target sees its reputation degrade, possibly facing congressional scrutiny or the prospect of not being reappointed. Reputation considerations of the central banker may not have first order welfare consequences for society if they are primarily about personal reputation of the central banker, or if the government finds it relatively easier to hire a new central banker once the current term ends.

3.4.1 Social Costs of Reputation

Although a personal reputational cost to the central banker may have negligible welfare consequences, the reputational cost may accrue not only to the central banker, but to the central bank itself. For example, a central bank that overshoots its target may suffer from a perception that it is not committed to its target, degrading confidence in future targets and de-anchoring inflation expectations. In this case, the reputational cost has first order welfare consequences.

We capture this possibility by assuming that central bank reputation, T , now appears in the true utility function, $U_t(\pi_t, E_t\pi_{t+1}, \theta_t, T_t)$. For illustrative purposes, let us consider a three-period variant of the model, where preferences are given by

$$E_0 \left[U_0(E\pi_1) + \beta U_1(\pi_1, E\pi_2, \theta_1, T_1) + \beta^2 U_2(\pi_2, \theta_2, T_2) \right]$$

where we normalize reputation to be $T_t \leq 0$. Suppose that reputation is separable in utility, so that we have $U_1 = u_1(\pi_1, E\pi_2, \theta_1) + T_1$ and similarly for U_2 . In this case, we can apply a standard procedure to characterize incentive compatible mechanisms using the Envelope Theorem. For example, at date 2 we have

$$T_2(\theta_1, \theta_2) = -u_2(\pi_2(\theta_1, \theta_2), \theta_2) + u_2(\pi_2(\theta_1, \underline{\theta}), \underline{\theta}) + T_2(\theta_1, \underline{\theta}) + \int_{\underline{\theta}}^{\theta_2} \frac{\partial u_2(\pi_2, s)}{\partial s} ds$$

which allows us to substitute out for T_2 in terms of the policies π_2 and the boundary value $T_2(\theta_1, \underline{\theta})$. Performing a similar procedure at date 1 and solving the relaxed social planning problem (not subject to monotonicity, we obtain the following result.

Proposition 5. *At interior values of (θ_1, θ_2) , inflation under the relaxed social planning problem is given by*

$$\underbrace{\left(1 - F(\theta_1)\right)}_{\text{"Information Rent"}} \frac{\partial^2 u_1}{\partial \pi_1 \partial \theta_1} = \left| \frac{\partial U_0}{\partial E\pi_1} \right| f(\theta_1) \quad (3.12)$$

$$\frac{\partial F(\theta_2|\theta_1)}{\partial \theta_1} \frac{\partial^2 u_2}{\partial \pi_2 \partial \theta_2} = \frac{1}{\beta} \frac{\partial^2 u_1}{\partial E\pi_2 \partial \theta_1} f(\theta_2|\theta_1) \quad (3.13)$$

Consider first equation (3.13), which determines inflation at date 1. The right hand side is the social cost of inflation at date 0 through the Phillips Curve, as in the constrained efficient

allocation. However, the left hand side now differs from the constrained efficient allocation rule. In particular, the left hand side is now the marginal effect of a change in π_1 on the *information rents* achieved by all types $\tilde{\theta}_1 \geq \theta_1$. The reason that the information rent, rather than utility, now appears is that the a decrease in reputation T_2 is required to maintain incentive compatibility when allowing higher- θ_1 types to adopt higher inflation. However, this decrease in reputation is now socially costly, which the government internalizes. As a result, it is only the information rents – that is, the additional utility gained by higher types by virtue of being higher types – that contribute on net to social welfare once the reputation costs have been accounted for.

This intuition carries through to the optimal choice of inflation at date 2. The optimal mechanism now trades off the benefit of date 2 inflation – increasing the date 2 information rent – against the cost of date 2 inflation – reducing the date 1 information rent. The cost of the Phillips Curve now reflects itself in effects on information rents.

Finally, we can consider the implications for the design of the function T . Under the same definition of v_2 as before, at date 2, local incentive compatibility implies that

$$\frac{\partial T_2}{\partial \theta_2} = -v_1 \frac{\partial \pi_2}{\partial \theta_2} + \underbrace{\left(v_1 - \frac{\partial u_2}{\partial \pi_2} \right) \frac{\partial \pi_2}{\partial \theta_2}}_{\text{Nonlinear Slope}}.$$

Although this is in fact the same general condition as before, the optimal allocation is no longer constrained efficient. As a result, the “inflation target” at date 1 now entails a nonlinear slope when mapping changes in inflation to changes in T_2 .

We can conduct a similar exercise at date 1. Suppose that we define v_1 as before. We can now write

$$\frac{\partial T_1}{\partial \theta_1} = -v_0 \frac{\partial \pi_1}{\partial \theta_1} + \underbrace{\left(v_0 - \frac{\partial u_1}{\partial \pi_1} \right) \frac{\partial \pi_1}{\partial \theta_1}}_{\text{Nonlinear Slope}} - \underbrace{\beta E \left[\frac{\partial T_2}{\partial \theta_1} - v_1 \pi_2 \frac{\partial f(\theta_2|\theta_1)}{\partial \theta_1} \middle| \theta_1 \right]}_{\text{Firm Expectations (“Date 2 Target”)}} - \underbrace{\beta E \left[\left(-v_1 + \frac{\partial u_2}{\partial \pi_2} \right) \frac{\partial \pi_2}{\partial \theta_1} \middle| \theta_1 \right]}_{\text{Deviation from Constrained Efficiency}}$$

The first three terms on the right hand side are familiar from before. The first term on the right hand side is the linear slope associated with implemented the constrained efficient allocation. The second term reflects nonlinearities in this slope.¹⁴ The third term is the direct impact of

¹⁴In the three period model, the linear slope is recovered at date 1 because there is no information rent at date 0.

the change in reported type on firm expectations, due to the change in perceived probability measure over the date 2 state. It gives rise to the target (intercept) in the baseline model. In other words, this term tells us that there is a term of the form $-v_1 E\pi_2$ in the date 2 transfer rule.

However, there is also a final term that reflects the extent to which the allocation under the optimal mechanism deviates from constrained efficiency. When the constrained efficient inflation at date 2 is not implemented, changes in the date 1 report impact central bank welfare to first order through changes in future inflation.

Relation to Section 3.3. The mechanism of Section 3.3 remains implementable in this framework, meaning that a dynamic inflation target can still be used to implement the constrained efficient allocation. However, this mechanism is no longer optimal as the reputation cost of the transfer now enters into social welfare.

Reputation Without Separability. To solve the problem above, we assumed reputation was separable in utility, which allowed for use of the same incentive compatibility constraints as before. Suppose that we do not have full separability, but we do have separability in θ_1 and T_1 , that is we can write $U_1 = u_1(\pi_1, E\pi_2, \theta_1) + v_1(\pi_1, E\pi_2, T_1)$. Notice that v_1 depends only on the reported type and on policy variables. In this case, we can define $V_1(\tilde{\theta}) = v_1(\pi_1, E\pi_2, T_1)$, with the constraint $V_1(\tilde{\theta}) \leq v_1(\pi_1(\tilde{\theta}), E\pi_2(\tilde{\theta}), 0)$ restricting maximum promisable utility. If under the optimal mechanism the maximum promisable utility constraint only binds at $\theta_1 = \underline{\theta}$, then we can conduct the analysis as before in the interior. Notice that there is an additional effect at the boundary of inflation expectations on maximum promisable utility, but not in the interior.

However, it is possible that maximum promisable utility may also bind above $\underline{\theta}$. In this case, we have a pooling region at low values of θ_1 of central banks with maximum reputation. Over this region, inflation expectations affect the central bank in part through this maximum promisable utility, generating a second impact, in addition to the forces from before.

However, notice that outside the pooling region, where the promisable utility constraint does not bind, the optimal allocation rule *does not* consider the impact of inflation expectations on welfare through v_1 . The intuition is that in this region, reputation and inflation expectations

are substitutes in enforcing incentive compatibility: a central bank can be made worse off either by having a lower reputation, or by facing higher inflation expectations. As a result, in the interior region higher expectations reduce the required loss of reputation.

3.5 Second-Best Mechanisms with Costly Transfers

In Section 3.2, we showed that a dynamic inflation target can implement the constrained efficient level of inflation, while Section 3.4 studies the case where T_t reflects a reputational cost that harms both the central bank and the government. In this section, we study the optimal mechanism when enforcement is costly in a conventional sense, for example associated with a monetary incentive scheme.

3.5.1 Model

We capture the social cost of implementing and enforcing a monetary policy mechanism by assuming that transfers are now costly to the government. Social preferences are now

$$\max \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t (U_t(\pi_t, \pi_{t+1}, \theta_t, \tilde{\theta}_t) - \kappa T_t) \right], \quad (3.14)$$

where $\kappa \geq 0$ captures the importance of this social cost.¹⁵

In previous sections, the costless nature of the mechanism implied participation constraints were irrelevant. We now introduce a date 0 participation constraint, given by

$$\mathcal{W}_0 \geq 0. \quad (3.15)$$

Without loss of generality, we have normalized the outside option in the participation constraint to 0.¹⁶ Recall that a mechanism is a mapping $(\pi_t, T_t) : \Theta^t \rightarrow \mathbb{R}^2$.

We now characterize local incentive compatibility from the Envelope Theorem, in order to solve the problem using a first-order approach that relaxes the required monotonicity

¹⁵This corresponds to a standard (quasilinear) transferable utility model. As usual, T_t may also correspond to non-quasilinear utilities, provided they are transferable in this form.

¹⁶If instead the outside option where \mathcal{W}_0 , the mechanism we derive still applies, except that the period-0 transfer would be adjusted by \mathcal{W}_0 for all types.

constraint.¹⁷ The Bellman representation of the central bank's problem is

$$\mathcal{W}_t(\theta^t) = \sup_{\tilde{\theta}} \{U_t(\pi_t, \pi_{t+1}, \theta_t, \tilde{\theta}_t) + T_t + \beta \mathbb{E}_t[\mathcal{W}_{t+1}|\theta_t]\} \quad (3.16)$$

As before, π_t, T_t are functions of the history of reported types, not true types, while \mathcal{W}_t is a function of both past *reported* types and the current *true* type.

We adopt the notational convention that $\frac{\partial U_t}{\partial \theta_t}$ is the derivative of U_t in the true type θ_t , while $\frac{\partial U_t}{\partial \tilde{\theta}_t}$ is the derivative in the reported type. In equilibrium, the reported type and the true type coincide, but the derivatives are distinct. Applying the Envelope Theorem and integrating over types, we obtain an integral incentive compatibility condition

$$\mathcal{W}_t(\theta^t) = \int_{\underline{\theta}}^{\theta_t} \frac{\partial U_t(\theta^{t-1}, s_t)}{\partial s_t} ds_t + \beta \int_{\underline{\theta}}^{\theta_t} \mathbb{E}_t \left[\mathcal{W}_{t+1} \frac{\partial f_t(\theta_{t+1}|s_t)/\partial s_t}{f_t(\theta_{t+1}|s_t)} \Big| s_t \right] ds_t \quad (3.17)$$

Integral incentive compatibility relates the total date- t utility to the central bank to two information rents. Note that due to shock persistence, the central bank earns information rents not only due to the effect on current flow utility, but also on the conditional probability distribution.¹⁸

Integral incentive compatibility (3.17) gives a Bellman representation to the value function, in terms of only the allocation rule. We can re-express this Bellman equation in sequence form by iterating the Bellman equation forward. Doing so, we obtain the following result characterizing this sequence representation.

Lemma 6. *The value function \mathcal{W}_t can be represented as*

$$\mathcal{W}_t(\theta^t) = \mathbb{E}_t \left[\sum_{s=0}^{\infty} \beta^s B_t^s(\theta^{t+s}) \Big| \theta_t \right] \quad \forall t,$$

where B_t^s is given by

$$B_t^s(\theta^{t+s}) = \prod_{k=0}^{s-1} \frac{1}{f_{t+k}(\theta_{t+k+1}|\theta_{t+k})} \int_{s_t \leq \theta_t, \dots, s_{t+s} \leq \theta_{t+s}} \frac{\partial U_{t+s}(\theta^{t-1}, s_t, \dots, s_{t+s})}{\partial s_{t+s}} \prod_{k=0}^{s-1} \frac{\partial f_{t+k}(\theta_{t+k+1}|s_{t+k})}{\partial s_{t+k}} ds_{t+s} \dots ds_t.$$

Lemma 6 allows us to represent the principal's optimization problem in a tractable way.

¹⁷See Pavan *et al.* (2014). We do not provide results on when the first order approach yields a globally incentive compatible mechanism.

¹⁸Without loss of generality, we have set the constant of integration $\mathcal{W}_t(\theta^{t-1}, \underline{\theta}) = 0$ and have used the outside option normalization $\underline{W}_0(\underline{\theta}) = 0$.

Given an allocation rule for inflation, we use the characterization of the value function in Lemma 6 as well as the Bellman equation to characterize the transfer rule which implements the allocation,

$$T_t = \mathcal{W}_t - U_t - \beta \mathbb{E}_t[\mathcal{W}_{t+1} | \theta_t].$$

We can then substitute the implementing taxes into the government's utility function, and obtain the following result characterizing the relaxed social planning problem.

Lemma 7. *The relaxed social planning problem can be written as*

$$\max_{\{\pi_t\}} \mathbb{E}_{-1} \left[\sum_{t=0}^{\infty} \beta^t \left[-\frac{\kappa}{1+\kappa} B_0^t + U_t \right] \right],$$

where B_0^t is given as in Lemma 6. The implementing transfer rule is given by

$$T_t = \mathcal{W}_t - U_t - \beta \mathbb{E}_t[\mathcal{W}_{t+1} | \theta_t],$$

where \mathcal{W}_t is given as a function of the allocation rule as in Lemma 6.

3.5.2 The Second-Best Allocation Rule

Lemma 7 provides a characterization of the relaxed social planning problem, subject to integral incentive compatibility. From this, we can characterize the optimal allocation.¹⁹

Proposition 8. *The solution to the optimal allocation rule is given by the first-order conditions*

$$\beta \frac{\partial U_t}{\partial \pi_t} f(\theta_t | \theta_{t-1}) = \overbrace{-\frac{\partial U_{t-1}}{\partial \pi_t}}^{\text{Time Consistency}} + \frac{\kappa}{1+\kappa} \underbrace{\left[\underbrace{\Gamma_{t-1}(\theta^{t-1}) \frac{\partial^2 U_{t-1}}{\partial \theta_{t-1} \partial \pi_t}}_{\text{Persistence + Time Consistency}} + \underbrace{\beta \Gamma_t(\theta^t) \frac{\partial^2 U_t}{\partial \theta_t \partial \pi_t(\theta^t)} f(\theta_t | \theta_{t-1})}_{\text{Persistence}} \right]}_{\text{Costly Transfers}} \quad \forall t \geq 1$$

¹⁹We characterize the optimal allocation assuming that π_t is interior.

$$\frac{\partial U_0}{\partial \pi_0} = \frac{\kappa}{1 + \kappa} \Gamma_0(\theta^0) \frac{\partial^2 U_0}{\partial \theta_0 \partial \pi_0}$$

where $\Gamma_t(\theta^t)$ is given by the recursive sequence

$$\Gamma_t(\theta^t) = \Gamma_{t-1}(\theta^{t-1}) \frac{1 - F(\theta_t | \theta_{t-1})}{f(\theta_t | \theta_{t-1})} \mathbb{E}_{t-1} \left[\frac{\partial f(s_t | \theta_{t-1}) / \partial \theta_{t-1}}{f(s_t | \theta_{t-1})} \middle| s_t \geq \theta_t \right] \quad (3.18)$$

with initial condition $\Gamma_0(\theta^0) = \frac{1 - F(\theta_0)}{f(\theta_0)}$.

There are three effects that cause the allocation rule to deviate from simple optimization of flow utility, $\frac{\partial U_t}{\partial \pi_t} = 0$, which would be the decision of the central bank if left uncontrolled by the government.

The first effect is a pure time consistency effect, reflected by the term $-\frac{\partial U_{t-1}}{\partial \pi_t}$. This effect arises because period- t inflation appears in period $t - 1$ flow utility, due to the Phillips curve relationship. This effect would exist even in the absence of persistent shocks and costly transfers. If transfers are costless or information is not persistent, then it is the only effect, and we return to the characterization in Section 3.2.

If transfers were not costly ($\kappa = 0$), the optimal allocation would be the constrained efficient allocation, and the dynamic inflation target would be the optimal mechanism. However, there are two effects arising from the fact that transfers are now costly. These effects reflect both the persistence of information and the time consistency problem. Because the central bank earns an information rent, implementing a given allocation rule is now costly. Suppose that on the margin, the government changes the allocation $\pi_t(\theta^t)$, holding all else fixed. This affects the required transfer at two points. First, it affects the required transfer in period t , which is related to the cross partial $\frac{\partial^2 U_t}{\partial \theta_t \partial \pi_t}$ owing to the Envelope Theorem characterization of transfers. Supposing that $\frac{\partial^2 U_t}{\partial \theta_t \partial \pi_t} > 0$ and that $\Gamma_t > 0$, then $\frac{\partial^2 U_t}{\partial \theta_t \partial \pi_t} > 0$ implies that an increase in π_t increases $\frac{\partial U_t}{\partial \theta_t} > 0$, which in turn increases the required transfer, generating a cost for the government. This is a standard effect that exists even without a time consistency problem.

However, changes in π_t also affect the required transfers at date $t - 1$, due to the Phillips curve relationship. Because π_t affects U_{t-1} , we have the analogous property applying. In particular if $\Gamma_{t-1} > 0$ and $\frac{\partial^2 U_{t-1}}{\partial \theta_{t-1} \partial \pi_t}$, then the effect is analogous. An increase in π_t causes an

increase in $\frac{\partial U_{t-1}}{\partial \theta_{t-1}}$, which increases the required transfer. In other words, this second effect is an interaction between the time consistency problem and the information persistence problem.

In either case, note that when the RHS is positive, we have $\frac{\partial U_{t-1}}{\partial \pi_t} + \beta \frac{\partial U_t}{\partial \pi_t} f(\theta_t | \theta_{t-1}) > 0$. This implies that from the social perspective (accounting for time consistency), it would be desirable to increase π_t further to reach the constrained efficient allocation. However, it is costly to provide the correct incentive scheme to do so, and as a result the time consistency problem is not fully corrected. Costly enforcement reduces the ability to combat the time consistency problem. Notice that this implies that the inflation target is *stricter* than the constrained efficient allocation.

Lastly, the initial condition $\Gamma_0(\theta^0) = \frac{1-F(\theta_0)}{f(\theta_0)}$ implies a conventional first-order condition for period-0 inflation

$$\frac{\partial U_0}{\partial \pi_0} - \frac{\kappa}{1+\kappa} \Gamma_0(\theta^0) \frac{\partial^2 U_0}{\partial \theta_0 \partial \pi_0} = 0$$

This reflects a standard virtual value relationship. The component $(1 - F(\theta_0))/f(\theta_0) > 0$ of virtual value is relevant in both period 1 and in period 0, due to the time consistency problem. Moreover, this value $(1 - F(\theta_0))/f(\theta_0) > 0$ is persistent over time through the evolution of Γ_t . In other words, if it is large initially, then Γ_t is also larger in magnitude along every history.

It should be noted that the constrained efficient allocation is still *implementable*, but is no longer optimal. In particular, at the constrained efficient allocation, the informational effects still exactly offset one another, and a dynamic inflation target still implements that allocation. However, at the constrained efficient allocation the marginal benefit of higher inflation is zero, while the marginal cost of the enforcement mechanism is not. As a result, the government no longer finds it optimal to implement the constrained efficient allocation that would have arisen under zero costs. This helps to understand why the optimal mechanism here deviates from a precise dynamic inflation target: once moved away from the constrained efficient allocation, the informational effect on firms no longer exactly offsets the informational effect on government transfers.

Nevertheless, the properties of the optimal mechanism still bear similarities to the dynamic inflation target. In particular, the time consistency term is still a constant given the Phillips curve relationship, generating a component of the optimal mechanism that resembles the dynamic

target.

3.5.3 Sufficient Statistics for the Optimal Mechanism

Although in principle we are now solving a dynamic mechanism design problem with persistence and a time consistency problem, the allocation rule derived in Proposition 8 implies a sufficient statistics result for characterizing the optimal allocation rule. In particular, at date t , we need to know two sufficient statistics for past information in order to derive the allocation rule.

The first sufficient statistic is the function $\Gamma_t(\theta^t)$, which itself has two sufficient statistics - the constants Γ_{t-1} and θ_{t-1} . This first pair of sufficient statistic summarizes the information relevant from the fact that shocks are persistent over time. It is separate from the time consistency problem, and would exist even if there were no time consistency problem. This sufficient statistic is irrelevant, however, in the absence of persistent information. In other words, if there is time consistency problems but idiosyncratic shocks, we do not need to keep track of this information.

The second sufficient statistic is a “wedge” ν_{t-1} , which is a date $t - 1$ adapted constant given by

$$\nu_{t-1} = -\frac{1}{\beta} \left[\frac{\partial U_{t-1}}{\partial \mathbb{E}_{t-1} \pi_t} - \frac{\kappa}{1 + \kappa} \Gamma_{t-1} \frac{\partial^2 U_{t-1}}{\partial \theta_{t-1} \partial \mathbb{E}_{t-1} \pi_t} \right] \quad (3.19)$$

ν_{t-1} is a sufficient statistic for the time consistency problem, including the interaction between the time consistency problem and the information persistence problem. ν_{t-1} would equal zero in absence of the time consistency problem, whether or not there were persistent shocks.

Using this triple $(\Gamma_{t-1}, \theta_{t-1}, \nu_{t-1})$, we can solve for the optimal allocation from date t onward. From there, we can use the central bank value function to back out the required transfer rule. As a result, this triple are sufficient statistics for the shock history θ^{t-1} under the optimal mechanism. This fact is summarized in the following corollary to Proposition 8.

Corollary 9. The triple $(\nu_{t-1}, \theta_{t-1}, \Gamma_{t-1})$ of date $t - 1$ adapted constants is a sufficient statistic for the history of shocks θ^{t-1} at date t . The allocation rule can be written as

$$\frac{\partial u_t}{\partial \pi_t} = \nu_{t-1} + \frac{\kappa}{1 + \kappa} \Gamma_t \frac{\partial^2 u_t}{\partial \theta_t \partial \pi_t} f(\theta_t | \theta_{t-1}) \quad \forall t \geq 1 \quad (3.20)$$

Corollary 9 further helps to understand how this mechanism resembles a dynamic inflation target. Over time, the time consistency term ν_{t-1} , which is a constant from a date t perspective, evolves in response to shocks. Smaller slopes imply lower inflation penalties and higher average inflation, whereas larger slopes imply larger penalties and lower inflation. In this sense, the slope ν_{t-1} now tells us information about the nature of the target.

3.5.4 Reversion to Dynamic Inflation Target

Proposition 8 tells us that the optimal mechanism reverts to a dynamic inflation target at both extremes of the shock distribution. In other words, there is both a no top distortion *and* a no bottom distortion result: if at period t we have $\theta_t = \bar{\theta}$ or $\theta_t = \underline{\theta}$, then $\Gamma_t = 0$, and hence $\Gamma_{t+s} = 0 \forall t \geq s$. This implies that the entire costly transfer term disappears, at which point the optimal mechanism reverts to the dynamic inflation target.

Corollary 10. If $\theta_t \in \{\underline{\theta}, \bar{\theta}\}$, then $\Gamma_{t+s} = 0 \forall s \geq 0$. ν_{t+s-1} is a sufficient statistic at date $t + s \forall s \geq 0$. The optimal allocation at dates $t + s$ ($s \geq 1$) is implemented by a dynamic inflation target (as in Proposition 4).

The no top distortion result closely resembles normal top distortion results in the absence of time consistency. The normal result states that after $\theta_t = \bar{\theta}$, we revert to the optimal allocation rule $\frac{\partial u_{t+s}}{\partial \pi_{t+s}} = 0 \forall s \geq 0$. In our model, the optimal allocation rule instead satisfies $\frac{\partial u_{t+s-1}}{\partial \pi_{t+s}} + \beta \frac{\partial u_{t+s}}{\partial \pi_{t+s}} f(\theta_{t+s} | \theta_{t+s-1}) = 0$, due to the Phillips curve. No top distortion reverts the economy to this efficient relationship. However, this efficient relationship is still distorted by the time consistency problem, when the central bank is left to set inflation in an uncontrolled manner. As a result, the wedge ν_{t+s} are still relevant, even after reaching a state $\theta_t = \bar{\theta}$. Moreover due to information persistence, we also have a no distortion at the bottom result.²⁰

As a result, not only does the optimal mechanism has a component that resembles the dynamic inflation target throughout the shock distribution, but it reverts fully to the dynamic inflation target at the limits of the distribution.

²⁰See Pavan *et al.* (2014) for related results.

3.5.5 Second best with Average Transfers

In the baseline model, we impose the assumption that the outside option takes the form $\mathcal{W}_0(\theta^0) \geq 0$. We might alternatively have expressed this in the form

$$\int_{\theta_0} \mathcal{W}_0(\theta^0) f(\theta_0 | \theta_{-1}) d\theta_0 \geq 0$$

The core difference between these two assumptions from a modeling perspective is on the timing of information arrival versus the participation decision. Under the baseline assumption, either θ_0 is already known to the central bank, or the central bank has the opportunity to revert to the outside option after learning θ_0 . Under the second assumption, θ_0 is not known to the central bank, and the central bank does not have the option to revert to the outside option after learning it.

Under this alternative structure, the optimality of the dynamic inflation target returns. In particular, implementable allocations are still defined as in Lemma 6, while the transfer rule is $T_t(\theta^t) = \mathcal{W}_t - U_t - \beta \mathbb{E}_t[\mathcal{W}_{t+1} | \theta_t]$. The average participation constraint implies that we have

$$0 = E_{-1} \mathcal{W}_0 = \mathbb{E}_{-1} \sum_{t=0}^{\infty} \beta^t (U_t + T_t),$$

which is markedly different from the baseline model. In particular, substituting this expression into social welfare, we obtain the social optimization problem

$$\max_{\{\pi_t\}} \mathbb{E}_{-1} \sum_{t=0}^{\infty} \beta^t (1 + \kappa) U_t$$

implying that the optimal allocation rule is constrained efficient. From here, we obtain the optimality of the dynamic inflation target.

Proposition 11. *Suppose that the participation constraint takes the form*

$$\int_{\theta_0} \mathcal{W}_0(\theta^0) f(\theta_0 | \theta_{-1}) d\theta_0 \geq 0$$

Then, the optimal mechanism is a dynamic inflation target, and yields the constrained efficient allocation.

The intuition behind Proposition 8 is straight-forward: under the average constraint, the government can capture the full social surplus and simply reduce the average transfer to the

central bank at date 0 to satisfy the participation constraint. This implies that the government chooses the mechanism and allocation that maximize social surplus, which is the dynamic inflation target.

3.5.6 Discussion

The results of this suggestion imply that even under costly enforcement (transfers), the optimal mechanism has properties that resemble a dynamic inflation target, in particular the time-varying slope ν_{t-1} associated with the time consistency problem. Moreover, the optimal mechanism reverts to the dynamic inflation target at the extremes of the shock distribution.

Taken together, if policymakers believe that the costs of enforcement are second-order next to the social welfare inflation-output trade-off, or believe that the target adjustment to (e.g.) 4% is a movement close to the upper limit of the distribution, then adopting a simple dynamic inflation target may be a good approximation to the true optimal mechanism. Moreover, the dynamic inflation target can always be used to implement the constrained efficient allocation, even if these statements are not true.

3.6 Conclusion

We provide a theory of dynamic inflation targets. We show that a dynamic inflation targeting mechanism can implement the constrained efficient level of inflation, and that when enforcement costs of using the mechanism are negligible the dynamic inflation target is the optimal mechanism. The key property of the dynamic inflation target is that it is updated *one period in advance*, which mitigates the time consistency problem in target setting. This suggests that the key property of target efficacy is advanced commitment. The results provide guidance on the mechanism by which to adjust an inflation target without exacerbating the underlying time consistency problem, suggesting that a mechanism of adjustment at restricted points in time, for example every five years as with the Bank of Canada, would be the desirable adjustment method.

References

- ABOWD, J. M. and ZELLNER, A. (1985). Estimating gross labor-force flows. *Journal of Business & Economic Statistics*, **3** (3), 254–283.
- ACHARYA, S. and DOGRA, K. (2020). Understanding HANK: Insights from a PRANK. *Econometrica*, **88** (3), 1113–1158.
- ACHDOU, Y., HAN, J., LASRY, J.-M., LIONS, P.-L. and MOLL, B. (2015). Heterogeneous agent models in continuous time.
- AHIR, H., BLOOM, N. and FURCERI, D. (2018). The world uncertainty index. *Available at SSRN* 3275033.
- AHN, S. (2019). Computing the distribution: Adaptive finite volume methods for economic models with heterogeneous agents.
- , KAPLAN, G., MOLL, B., WINBERRY, T. and WOLF, C. (2017). Micro Heterogeneity and Aggregate Consumption Dynamics. *NBER Macroeconomics Annual*, *forthcoming*.
- AIYAGARI, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, **109** (3), 659–684.
- ALGAN, Y., ALLAIS, O. and DEN HAAN, W. J. (2008). Solving heterogeneous-agent models with parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, **32** (3), 875–908.
- , —, DEN HAAN, W. J. and RENDAHL, P. (2014). Solving and simulating models with heterogeneous agents and aggregate uncertainty. In *Handbook of computational economics*, vol. 3, Elsevier, pp. 277–324.
- ALTIG, D., CHRISTIANO, L. J., EICHENBAUM, M. and LINDE, J. (2011). Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic dynamics*, **14** (2), 225–247.
- AMADOR, M., WERNING, I. and ANGELETOS, G. M. (2006). Commitment vs. flexibility. *Econometrica*, **74** (2), 365–396.
- ARCIDIACONO, P., BAYER, P., BLEVINS, J. R. and ELLICKSON, P. B. (2016). Estimation of dynamic discrete choice models in continuous time with an application to retail competition. *Review of Economic Studies*, **83** (3), 889–931.
- ARELLANO, C., BAI, Y. and KEHOE, P. J. (2019). Financial frictions and fluctuations in volatility. *Journal of Political Economy*, **127** (5), 2049–2103.

- ATHEY, S., ATKESON, A. and KEHOE, P. J. (2005). The optimal degree of discretion in monetary policy. *Econometrica*, **73** (5), 1431–1475.
- AUCLERT, A., BARDÓCZY, B., ROGNLIE, M. and STRAUB, L. (2019). *Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models*. Tech. rep., National Bureau of Economic Research.
- , ROGNLIE, M. and STRAUB, L. (2018). *The intertemporal keynesian cross*. Tech. rep., National Bureau of Economic Research.
- , — and — (2020). *Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model*. Tech. rep., National Bureau of Economic Research.
- BACHMANN, R., MOSCARINI, G. and OTHERS (2011). Business cycles and endogenous uncertainty. Citeseer.
- BAJARI, P., BENKARD, C. L. and LEVIN, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, **75** (5), 1331–1370.
- BAKER, S. R., BLOOM, N. and DAVIS, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, **131** (4), 1593–1636.
- BALL, L., GAGNON, J., HONOHAN, P. and SIGNE, K. (2016). What Else Can Central Banks Do? Geneva Report on the World Economy no. 18. London: Centre for Economic Policy Research. (August).
- BARLES, G. and SOUGANIDIS, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic analysis*, **4** (3), 271–283.
- BARRO, R. J. and GORDON, D. B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, **12** (1), 101–121.
- BASU, S. and BUNDICK, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, **85** (3), 937–958.
- BAYER, C., LÜTTICKE, R., PHAM-DAO, L. and TJADEN, V. (2019). Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk. *Econometrica*, **87** (1), 255–290.
- BEAUDRY, P. and PAGES, C. (2001). The cost of business cycles and the stabilization value of unemployment insurance. *European Economic Review*, **45** (8), 1545–1572.
- BELLMAN, R. E. (1961). *Adaptive control processes*.
- BERGER, D. and VAVRA, J. (2019). Shocks versus Responsiveness: What Drives Time-Varying Dispersion? *Journal of Political Economy*, **127** (5), 2104–2142.
- BERNANKE, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *The quarterly journal of economics*, **98** (1), 85–106.
- BESHEARS, J., CHOI, J., CLAYTON, C., HARRIS, C., LAIBSON, D. and MADRIAN, B. (2020). Optimal Illiquidity.

- BILBIIE, F. O. (2020). The new Keynesian cross. *Journal of Monetary Economics*, **114**, 90–108.
- BLANCHARD, O. J. (1985). Debt, deficits, and finite horizons. *Journal of political economy*, **93** (2), 223–247.
- , DIAMOND, P., HALL, R. E. and MURPHY, K. (1990). The cyclical behavior of the gross flows of US workers. *Brookings Papers on Economic Activity*, **1990** (2), 85–155.
- BLEVINS, J. R. (2018). *Identification and Estimation of Continuous Time Dynamic Discrete Choice Games*. Tech. rep.
- BLOOM, N. (2009). The impact of uncertainty shocks. *econometrica*, **77** (3), 623–685.
- (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, **28** (2), 153–176.
- , FLOETOTTO, M., JAIMOVICH, N., SAPORTA-EKSTEN, I. and TERRY, S. J. (2018). Really uncertain business cycles. *Econometrica*, **86** (3), 1031–1065.
- BOKANOWSKI, O., GARCKE, J., GRIEBEL, M. and KLOMPIAKER, I. (2013). An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. *Journal of Scientific Computing*, **55** (3), 575–605.
- BOPPART, T., KRUSELL, P. and MITMAN, K. (2018). Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative. *Journal of Economic Dynamics and Control*, **89**, 68–92.
- BRUMM, J. and SCHEIDEGGER, S. (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica*, **85** (5), 1575–1612.
- BRUNNERMEIER, M. K. and SANNIKOV, Y. (2014). A macroeconomic model with a financial sector. *The American Economic Review*, **104** (2), 379–421.
- BUNGARTZ, H.-J. and GRIEBEL, M. (2004). Sparse grids. *Acta numerica*, **13** (1), 147–269.
- CANZONERI, M. B. (1985). Monetary policy games and the role of private information. *American Economic Review*, **75** (5), 1056–1070.
- CAPLIN, A. and NALEBUFF, B. (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, **59** (1), 25–59.
- CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. and LIONS, P.-L. (2015). The master equation and the convergence problem in mean field games. *arXiv preprint arXiv:1509.02505*.
- CHALLE, E., MATHERON, J., RAGOT, X. and RUBIO-RAMIREZ, J. F. (2017). Precautionary saving and aggregate demand. *Quantitative Economics*, **8** (2), 435–478.
- CHODOROW-REICH, G. and KARABARBOUNIS, L. (2016). The cyclicity of the opportunity cost of employment. *Journal of Political Economy*, **124** (6), 1563–1618.
- CHRISTIANO, L., EICHENBAUM, M. and REBELO, S. (2011). When is the government spending multiplier large? *Journal of Political Economy*, **119** (1), 78–121.

- CHRISTIANO, L. J., EICHENBAUM, M. and EVANS, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, **113** (1), 1–45.
- DARBY, M. R., HALTIWANGER, J. C. and PLANT, M. W. (1985). *Unemployment-rate dynamics and persistent unemployment under rational expectations*. Tech. rep., National Bureau of Economic Research.
- , — and — (1986). *The ins and outs of unemployment: The ins win*. Tech. rep., National Bureau of Economic Research.
- DEN HAAN, W. J. (1996). Heterogeneity, aggregate uncertainty, and the short-term interest rate. *Journal of Business & Economic Statistics*, **14** (4), 399–411.
- DEN HAAN, W. J. (2010). Comparison of solutions to the incomplete markets model with aggregate uncertainty. *Journal of Economic Dynamics and Control*, **34** (1), 4–27.
- DEN HAAN, W. J. and OTHERS (1997). Solving dynamic models with aggregate shocks and heterogeneous agents. *Macroeconomic dynamics*, **1**, 355–386.
- , RENDAHL, P. and RIEGLER, M. (2018). Unemployment (fears) and deflationary spirals. *Journal of the European Economic Association*, **16** (5), 1281–1349.
- DORASZELSKI, U. and JUDD, K. L. (2012). Avoiding the curse of dimensionality in dynamic stochastic games. *Quantitative Economics*, **3** (1), 53–93.
- and SATTERTHWAITE, M. (2010). Computable markov-perfect industry dynamics. *The RAND Journal of Economics*, **41** (2), 215–243.
- DOU, W., FANG, X., LO, A. W. and UHLIG, H. (2020). Macro-finance models with nonlinear dynamics. *University of Chicago, Becker Friedman Institute for Economics Working Paper, The Rodney L. White Center Working Papers Series at the Wharton School*.
- DUARTE, V. (2018). Machine Learning for Continuous-Time Economics. *Available at SSRN 3012602*.
- EGGERTSSON, G. B. and KRUGMAN, P. (2012). Debt, deleveraging, and the liquidity trap: A Fisher-Minsky-Koo approach. *The Quarterly Journal of Economics*, **127** (3), 1469–1513.
- and WOODFORD, M. (2003). *Optimal monetary policy in a liquidity trap*. Tech. rep., National Bureau of Economic Research.
- ELSBY, M. W. L., MICHAELS, R. and SOLON, G. (2009). The Ins and Outs of Cyclical Unemployment. *American Economic Journal: Macroeconomics*, **1** (1), 84–110.
- ERCEG, C. J., HENDERSON, D. W. and LEVIN, A. T. (2000). Optimal monetary policy with staggered wage and price contracts. *Journal of monetary Economics*, **46** (2), 281–313.
- ERICSON, R. and PAKES, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, **62** (1), 53–82.
- FAJGELBAUM, P. D., SCHAAL, E. and TASCHEREAU-DUMOUCHEL, M. (2017). Uncertainty traps. *The Quarterly Journal of Economics*, **132** (4), 1641–1692.

- FERNÁNDEZ-VILLAVERDE, J., GORDON, G., GUERRÓN-QUINTANA, P. and RUBIO-RAMÍREZ, J. F. (2015a). Nonlinear adventures at the zero lower bound. *Journal of Economic Dynamics and Control*, **57**, 182–204.
- , GUERRÓN-QUINTANA, P., KUESTER, K. and RUBIO-RAMÍREZ, J. (2015b). Fiscal Volatility Shocks and Economic Activity. *American Economic Review*, **105** (11), 3352–3384.
- , HURTADO, S. and NUNO, G. (2019). *Financial frictions and the wealth distribution*. Tech. rep., National Bureau of Economic Research.
- FUJITA, S. and RAMEY, G. (2009). The cyclicity of separation and job finding rates. *International Economic Review*, **50** (2), 415–430.
- GALI, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.
- GARCKE, J. and KRÖNER, A. (2017). Suboptimal feedback control of PDEs by solving HJB equations on adaptive sparse grids. *Journal of Scientific Computing*, **70** (1), 1–28.
- and RUTTSCHIEDT, S. (2019). Finite Differences on Sparse Grids for Continuous Time Heterogeneous Agent Models.
- GERSTNER, T. and GRIEBEL, M. (2003). Dimension-adaptive tensor-product quadrature. *Computing*, **71** (1), 65–87.
- GERTLER, M. and KARADI, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, **58** (1), 17–34.
- and KIYOTAKI, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics*, vol. 3, Elsevier, pp. 547–599.
- GILCHRIST, S., SIM, J. W. and ZAKRAJŠEK, E. (2014). *Uncertainty, financial frictions, and investment dynamics*. Tech. rep., National Bureau of Economic Research.
- GRIEBEL, M. (1998). Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, **61** (2), 151–179.
- GUVENEN, F., OZKAN, S. and SONG, J. (2014). The nature of countercyclical income risk. *Journal of Political Economy*, **122** (3), 621–660.
- HALAC, M. and YARED, P. (2014). Fiscal Rules and Discretion Under Persistent Shocks. *Econometrica*, **82** (5), 1557–1614.
- and — (2019). Instrument-Based vs. Target-Based Rules.
- HE, Z. and KRISHNAMURTHY, A. (2013). Intermediary asset pricing. *The American Economic Review*, **103** (2), 732–770.
- HOTZ, V. J. and MILLER, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, **60** (3), 497–529.
- , —, SANDERS, S. and SMITH, J. (1994). A simulation estimator for dynamic models of discrete choice. *Review of Economic Studies*, **61** (2), 265–289.

- HUGGETT, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, **17** (5), 953–969.
- JUDD, K. L., MALIAR, L., MALIAR, S. and VALERO, R. (2014). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control*, **44**, 92–123.
- JURADO, K., LUDVIGSON, S. C. and NG, S. (2015). Measuring uncertainty. *American Economic Review*, **105** (3), 1177–1216.
- KAPLAN, G., MOLL, B. and VIOLANTE, G. L. (2018). Monetary policy according to HANK. *American Economic Review*, **108** (3), 697–743.
- KEHRIG, M. (2015). The cyclical nature of the productivity distribution. *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*.
- KEYNES, J. M. (1936). *The general theory of employment, interest, and money*. Springer.
- KIMBALL, M. S. (1990). Precautionary Saving in the Small and in the Large. *Econometrica*, **58** (1), 53–73.
- KLIMKE, A. and WOHLMUTH, B. (2005). Algorithm 847: spinterp: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB. *ACM Transactions on Mathematical Software (TOMS)*, **31** (4), 561–579.
- KREBS, T. (2007). Job displacement risk and the cost of business cycles. *American Economic Review*, **97** (3), 664–686.
- KRUEGER, D. and KUBLER, F. (2004). Computing equilibrium in OLG models with stochastic production. *Journal of Economic Dynamics and Control*, **28** (7), 1411–1436.
- KRUSELL, P., MUKOYAMA, T. and \CSAHIN, A. (2010). Labour-market matching with precautionary savings and aggregate fluctuations. *The Review of Economic Studies*, **77** (4), 1477–1507.
- and SMITH, A. A. J. (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy*, **106** (5), 867–896.
- KYDLAND, F. E. and PRESCOTT, E. C. (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, **85** (3), 473–491.
- LASRY, J.-M. and LIONS, P.-L. (2007). Mean field games. *Japanese journal of mathematics*, **2** (1), 229–260.
- LAUBACH, T. and WILLIAMS, J. C. (2003). Measuring the natural rate of interest. *Review of Economics and Statistics*, **85** (4), 1063–1070.
- and — (2016). Measuring the natural rate of interest redux. *Business Economics*, **51** (2), 57–67.
- LEDUC, S. and LIU, Z. (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, **82**, 20–35.
- LEVEQUE, R. J. (2007). *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*. SIAM.

- L'HUILLIER, J.-P. and SCHOENLE, R. (2019). Raising the Target: How Much Extra Room Does It Really Give?
- LIONS, P.-L. (2011). Mean field games and applications. In *Cours au College de France*, Springer.
- LORENZONI, G. (2008). Inefficient credit booms. *Review of Economic Studies*, **75** (3), 809–833.
- LUCAS, R. E. (1987). *Models of business cycles*, vol. 26. Basil Blackwell Oxford.
- (2003). Macroeconomic priorities. *American economic review*, **93** (1), 1–14.
- LUDVIGSON, S. C., MA, S. and NG, S. (2015). *Uncertainty and business cycles: exogenous impulse or endogenous response?* Tech. rep., National Bureau of Economic Research.
- MA, X. and ZABARAS, N. (2009). An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, **228** (8), 3084–3113.
- MARSTON, S. T., FELDSTEIN, M. and HYMANS, S. H. (1976). Employment instability and high unemployment rates. *Brookings Papers on Economic Activity*, **1976** (1), 169–210.
- MCFADDEN, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pp. 105–142.
- MCKAY, A. (2017). Time-varying idiosyncratic risk and aggregate consumption dynamics. *Journal of Monetary Economics*, **88**, 1–14.
- , NAKAMURA, E. and STEINSSON, J. (2016). The power of forward guidance revisited. *American Economic Review*, **106** (10), 3133–3158.
- and REIS, R. (2016). The role of automatic stabilizers in the US business cycle. *Econometrica*, **84** (1), 141–194.
- NAKATA, T. (2017). Uncertainty at the zero lower bound. *American Economic Journal: Macroeconomics*, **9** (3), 186–221.
- OH, H. and REIS, R. (2012). Targeted transfers and the fiscal response to the great recession. *Journal of Monetary Economics*, **59**, S50—S64.
- OKSENDAL, B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.
- OTTONELLO, P. and WINBERRY, T. (2017). Financial Heterogeneity and the Investment Channel of Monetary Policy.
- PAKES, A. and MCGUIRE, P. (1994). Computing markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics*, **25** (4), 555–589.
- and — (2001). Stochastic algorithms, symmetric markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica*, **69** (5), 1261–1281.

- , OSTROVSKY, M. and BERRY, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, **38** (2), 373–399.
- PÁSTOR, L. and VERONESI, P. (2013). Political uncertainty and risk premia. *Journal of financial Economics*, **110** (3), 520–545.
- PATTERSON, C. (2019). The matching multiplier and the amplification of recessions. In *2019 Meeting Papers*, Society for Economic Dynamics, vol. 95.
- PAVAN, A., SEGAL, I. and TOIKKA, J. (2014). Dynamic Mechanism Design: A Myersonian Approach. *Econometrica*, **82** (2), 601–653.
- PERSSON, T. and TABELLINI, G. (1993). Designing institutions for monetary stability. *Carnegie-Rochester Confer. Series on Public Policy*, **39** (C), 53–84.
- PLANTE, M., RICHTER, A. W. and THROCKMORTON, N. A. (2018). The zero lower bound and endogenous uncertainty. *The Economic Journal*, **128** (611), 1730–1757.
- POTERBA, J. M. and SUMMERS, L. H. (1986). Reporting errors and labor market dynamics. *Econometrica: Journal of the Econometric Society*, pp. 1319–1338.
- POWELL, J. H. (2019). Monetary policy in the post-crisis era. Speech at "Bretton Woods: 75 Years Later - Thinking about the Next 75".
- PRÖHL, E. (2019). Approximating equilibria with ex-post heterogeneity and aggregate risk. *Swiss Finance Institute Research Paper*, (17-63).
- RAVN, M. O. and STERK, V. (2016). Macroeconomic fluctuations with HANK & SAM: An analytical approach. *Journal of the European Economic Association*.
- and — (2017). Job uncertainty and deep recessions. *Journal of Monetary Economics*, **90**, 125–141.
- REITER, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, **33** (3), 649–665.
- (2010). Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control*, **34** (1), 28–35.
- ROGOFF, K. (1985). The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics*, **100** (4), 1169–1189.
- ROTEMBERG, J. J. (1982). Sticky prices in the United States. *Journal of Political Economy*, **90** (6), 1187–1211.
- RUTTSCHIEDT, S. (2018). Adaptive Sparse Grids for Solving Continuous Time Heterogeneous Agent Models.
- SCHAAB, A. (2020). Micro and Macro Uncertainty.
- and ZHANG, A. (2020). Using Adaptive Sparse Grids to Solve Heterogeneous-Agent Models in Continuous Time.

- SCHIEKOFER, T. (1998). *Die Methode der Finiten Differenzen auf d unnen Gittern zur L osung elliptischer und parabolischer partieller Di erentialgleichungen*. Ph.D. thesis, PhD thesis, Universit at Bonn.
- and ZUMBUSCH, G. (1998). Software concepts of a sparse grid finite difference code. *Concepts of Numerical Software, Notes in Numerical Fluid Mechanics*. Vieweg, Braunschweig.
- SCHMIDT, L. (2016). Climbing and falling off the ladder: Asset pricing implications of labor market event risk. *Available at SSRN 2471342*.
- SCHMITT-GROHÉ, S. and URIBE, M. (2005). Optimal fiscal and monetary policy in a medium-scale macroeconomic model. *NBER Macroeconomics Annual*, **20**, 383–425.
- SHIMER, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, **95** (1), 25–49.
- (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, **15** (2), 127–148.
- SMOLYAK, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Doklady Akademii Nauk*, Russian Academy of Sciences, vol. 148, pp. 1042–1045.
- STORESLETTEN, K., TELMER, C. I. and YARON, A. (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of political Economy*, **112** (3), 695–717.
- STRAUB, L. and ULBRICHT, R. (2017). Endogenous uncertainty and credit crunches. *Available at SSRN 2668078*.
- VAN NIEUWERBURGH, S. and VELDKAMP, L. (2006). Learning asymmetries in real business cycles. *Journal of monetary Economics*, **53** (4), 753–772.
- WALSH, C. E. (1995). Optimal contracts for central bankers. *American Economic Review*, **85** (1), 150–167.
- WERNING, I. (2011). *Managing a liquidity trap: Monetary and fiscal policy*. Tech. rep., National Bureau of Economic Research.
- (2015). *Incomplete markets and aggregate demand*. Tech. rep., National Bureau of Economic Research.
- WINBERRY, T. (2020). Lumpy Investment, Business Cycles, and Stimulus Policy.
- YOUNG, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, **34** (1), 36–41.
- ZARIPHOUPOULOU, T. (1994). Consumption-investment models with constraints. *SIAM Journal on Control and Optimization*, **32** (1), 59–85.
- ZENGER, C. (1991). Sparse grids. In *Proceedings of the Research Workshop of the Israel Science Foundation on Multiscale Phenomenon, Modelling and Computation*, p. 86.
- ZUMBUSCH, G. W. (2000). A sparse grid PDE solver; discretization, adaptivity, software design and parallelization. In *Advances in Software Tools for Scientific Computing*, Springer, pp. 133–177.

Appendix A

Appendix to Chapter 1

A.1 Appendix for Section 1.3: Quantitative Model

The material in this section takes as given that the model admits a finite-dimensional aggregate state space representation. This will of course be true for each element in the sequence of approximating economies specified by my solution method. For additional details, see Section 1.4 in the main text or Appendix A.3.

I will formalize this assumption before I proceed. In particular, as in the main text, I denote the aggregate state of the economy at time t by $\Gamma_t \in \mathbb{R}^N$ for some N . In practice, for the baseline model we will have $\Gamma_t = (\rho_t, \alpha_t)$, where α_t parameterizes the finite-dimensional distribution representation of my solution method. However, for this section it will suffice to postulate an arbitrary aggregate state vector Γ_t which follows a (Markov) diffusion so that

$$d\Gamma_t = \mu_\Gamma(\Gamma_t)dt + \sigma_\Gamma(\Gamma_t)dB,$$

where B_t is a one-dimensional standard Brownian motion, representing the sole aggregate risk factor of my model, and

$$\mu_\Gamma : \mathbb{R}^N \rightarrow \mathbb{R}^N, \quad \sigma_\Gamma : \mathbb{R}^N \rightarrow \mathbb{R}^N.$$

Formally, Γ_t is a *time-homogeneous Ito diffusion* and I assume that it satisfies the usual regularity conditions (see e.g. Oksendal (2013)). In other words, the coefficients μ_Γ and σ_Γ depend on time t only *through* Γ_t , and the process Γ_t therefore satisfies the Markov property for Ito diffusions.

Before proceeding, I will introduce the notion of *infinitesimal generator* which is a composite functional operator associated with a particular diffusion process. This object and its properties are at the heart of Proposition 3 as I already anticipated in the main text.

Definition 12. The infinitesimal generator of a time-homogeneous Ito diffusion process X_t in \mathbb{R}^N is defined as

$$\mathcal{A}_X f(x) = \lim_{t \rightarrow 0} \frac{\mathbb{E}_t[f(X_t)] - f(x)}{t}$$

where $X_0 = x \in \mathbb{R}^n$ and $f : \mathbb{R}^N \rightarrow \mathbb{R}$

In particular, the generator of the diffusion process Γ_t is then given by

$$\mathcal{A}_\Gamma f(x) = \mu_\Gamma(x) D_x f + \frac{1}{2} \sigma_\Gamma^T(x) D_{xx} f \sigma_\Gamma(x)$$

where the T superscript denotes a transpose. To arrive at this form, the assumption of some regularity conditions on f are necessary. The notation of (partial) derivative operators is such that $D_x f$ denotes the Jacobian of f and $D_{xx} f$ its Hessian matrix.

A.1.1 Household Problem

In the quantitative model, households consume, save, work, and make portfolio allocation decisions. The problem of a household is to maximize the objective function (1) in the main text subject to the set of budget, borrowing and short-sale constraints, and taking as given (i) all aggregate prices, (ii) the economy's aggregate law of motion, $d\Gamma_t$, and (iii) the idiosyncratic employment transition process.

It is easiest to work with a *recursive* representation of the household problem, which is possible under the assumption of a (time-homogeneous) aggregate state space representation (see above).¹ The appropriate state space on which we can formulate such a recursive representation is, of course, (a, k, z, Γ) where (a, k, z) describes the household's idiosyncratic state given by liquid and illiquid asset holdings, as well as employment status. As in the main text, I use

¹I will formally discuss the recursive representation of the household problem when Γ_t is *infinite-dimensional* in Appendix A.3. The discussion here is understood to refer to an *approximate economy* under a finite-dimensional distribution representation.

superscript j notation to refer to a household with employment status z^j , for $j \in \{U, E\}$. In sum, the household consumption *policy function*, which I will characterize below, is then given by $c^j(a, k, \Gamma)$ for a household in employment state j , with asset holdings (a, k) , and in the aggregate state Γ .

Derivation of the household budget constraint

In the main text, I work directly with *real* household asset holdings. In this subsection, I start from the nominal household budget constraint and derive its real analog, deflated by the household consumer price index. Recall that, in the *state space* representation of the problem, all aggregate prices are functions of the aggregate state of the economy, so that for example the real interest rate is given by $r_t = r(\Gamma_t)$. I will notationally suppress this dependence on Γ whenever there is no ambiguity. Similarly, I will suppress the dependence of functions on the idiosyncratic state (a, k, z) for ease of exposition.

Nominal consumption outlays of the household (in employment state j) are given by Pc^j , where P denotes the price index of the consumption bundle (CPI). I define a as the stock of “bond” holdings (liquid asset) and, similarly, k as the stock of “capital” holdings (illiquid asset). As such, the household budget constraint is given by an equation for the evolution of *liquid wealth*, Pa . We have

$$d(Pa) = \left[(i + \zeta)Pa + \left(i^k + \frac{P\Pi^Q}{K} \right)k + Pe - Q\iota^j - P\psi(\iota, k) - Pc \right] dt.$$

Several observations are in order. The portfolio returns are expressed in nominal terms here, where i is the nominal interest rate, Pa is the nominal bond value on which the nominal interest rate accrues (and on which annuity payments ζ are made). Similarly, i^k is the nominal rental rate per unit of capital k and $P\Pi^Q$ denotes the aggregate nominal profits of the capital producing sector. Q is the price of capital goods, where ι denotes the rate at which households purchase capital goods from the capital producer. I assume that the adjustment cost is paid in units of the consumption good, which is why it is valued at P . Finally, Pe denotes nominal household earnings.

CPI price inflation is defined as dP/P . The key question, then, is whether dP is stochastic,

in the sense that it loads directly on the aggregate risk factor dB . As I will discuss in more detail in Appendix A.1.3, dP has no diffusion term as long as we assume price stickiness (under Rotemberg adjustment costs). In this case, I can define CPI inflation as

$$\frac{dP}{P} = \pi dt.$$

Returning to the nominal budget constraint and using the fact that dP has no diffusion term, we have $d(Pa) = Pda + adP$ and, dividing by P and rearranging, I arrive at

$$\frac{da}{dt} = (i - \pi + \zeta)a + \left(\frac{i^k}{P} + \frac{\Pi Q}{K} \right) k + e - q\iota - \psi(\iota, k) - c,$$

where $r = i - \pi$ is the real riskfree rate, $r^k = i^k/P$ is the real rental rate, and $q = Q/P$ is the relative price of capital goods.

Household net worth

The definition of household net worth is conceptually complicated in this setting by the presence of adjustment costs. Similarly, due to these adjustment costs, it will generically not be possible to reduce the state space of the household from (a, k) to a single net worth variable. In other words, household behavior generally depends not only on wealth but rather on the portfolio split between liquid and illiquid assets.

Nonetheless, it is useful to consider how household net worth may be specified. I will define net worth n as though there were no adjustment costs, so that $n = a + qk$. I have solved a version of the model where there are indeed no adjustment cost at the micro level, so that the household portfolio choice problem can be expressed in terms of the single state variable n , as well as a portfolio share θ of risky (but no longer illiquid) assets. Such setting lends itself straightforwardly to a study of consumption-based asset pricing with heterogeneous households. Details are available upon request.

While such a state space reduction is not possible here, it is still useful to characterize the law of motion of household net worth as defined above. To do so, I specify a diffusion process

for the relative price of capital given by

$$\frac{dq}{q} = \mu_Q dt + \sigma_Q dB,$$

where μ_Q and σ_Q are equilibrium objects. Also recall that the household's illiquid asset position evolves non-stochastically as in $dk = [(\zeta - \delta)k + \iota]dt$.² Net worth therefore evolves as $dn = da + kdq + qdk$, or

$$\frac{1}{dt}dn = (r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K}\right)k + e - q\iota - \psi(\iota, k) - c + kq\mu_Q + kq\sigma_Q \frac{dB}{dt} + q(\zeta - \delta)k + q\iota.$$

Simplifying and rearranging, we arrive at

$$dn = \left[(r + \zeta)n + e - \psi(\iota, k) - c \right] dt + \underbrace{\left(\frac{r^k}{q} + \frac{\Pi^Q}{qK} \right) qk}_{\text{Dividend yield}} dt + \underbrace{\left[(\mu_Q - \delta - r) dt + \sigma_Q dB \right]}_{\text{Excess capital gains rate}} qk$$

This derivation highlights that household net worth indeed loads on the aggregate risk factor dB via a capital gains term, and thus follows a stochastic differential equation. On the other hand, the *stocks* of liquid and illiquid assets held by the household, a and k , are not stochastic in this sense. In other words, it is the *value* of these positions that evolves stochastically but not their stocks, whose evolution is controlled by the household.

Household value function and FOCs

This subsection derives the optimality conditions for household behavior. Recognizing that labor supply is set by unions, each household is left with two active choice variables, consumption c and investment ι , i.e. the purchase of the illiquid asset.

The formal statement of this problem for a household that starts in state (a, k, z^j, Γ) at time $t = 0$ is

$$V^j(a, k, z, \Gamma) = \max_{c, h, \iota} \mathbb{E}_0 \left[\int_0^\infty e^{-(\rho(\Gamma) + \zeta)t} u(c, h) dt \right]$$

²In the presence of capital quality shocks as in Brunnermeier and Sannikov (2014), for example, dk would instead be a stochastic differential equation, and a covariance term $(dk)(dq)$ would emerge.

subject to

$$\dot{a} = \left(r(\Gamma) + \zeta \right) a + \left(r^k(\Gamma) + \frac{\Pi^Q(\Gamma)}{K(\Gamma)} \right) k + e^j(a, k, \Gamma) - q(\Gamma)\iota - \psi(\iota, k) - c$$

$$\dot{k} = (\zeta - \delta)k + \iota$$

$$h = H(\Gamma)$$

$$dz^j = (z^{-j} - z^j) dN\left(\lambda^j(a, k, \Gamma)\right)$$

$$d\Gamma = \mu_\Gamma(\Gamma)dt + \sigma_\Gamma(\Gamma)dB$$

$$a \geq \underline{a} \quad \text{and} \quad k \geq 0,$$

taking as given all aggregate prices (as functions of the aggregate state Γ). The constraint $h = H(\Gamma)$ represents the assumption that all labor supply is set symmetrically by labor unions, and households take their quota of work hours as given. Household earnings are given by

$$e^j(a, k, \Gamma) = e^j(\Gamma) = (1 - \tau^{\text{lab}})z^j w(\Gamma)H(\Gamma) + \tau^{\text{lump}}(\Gamma) + \tau^{\text{UI}}(z^j).$$

Finally, as discussed in the main text, $dN(\cdot)$ represents a continuous-time Poisson process that is uncorrelated across households, and which takes as its argument the associated transition rate. In our case, this employment transition rate $\lambda^j(a, k, \Gamma)$ is given as an arbitrary function on the household state space.

Recursive representation. It is easiest to work with a recursive representation of the household problem. It is cast in terms of a partial differential equation (PDE) that represents the continuous-time analog of the discrete-time Bellman equation, called the Hamilton-Jacobi-Bellman (HJB) equation.

The HJB equation associated with the household problem (under the assumption that Γ is

finite-dimensional) is given by

$$\begin{aligned}
(\rho + \zeta)V^j = \max_{c^j, \iota^j} & \left\{ u(c^j, H) + V_k^j [(\zeta - \delta)k + \iota^j] \right. \\
& \left. + V_a^j \left[(r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K} \right) k + e^j - q\iota^j - \psi(\iota^j, k) - c^j \right] \right\} \\
& + \lambda^j [V^{-j} - V^j] + \mu_\Gamma V_\Gamma + \frac{1}{2} \sigma_\Gamma^T V_{\Gamma\Gamma} \sigma_\Gamma
\end{aligned}$$

where I again suppress the dependence of objects on state variables. I only explicitly account for the dependence on employment status using j superscripts in order to highlight that, really, this is a system of PDEs, one associated with each j . I use the shorthand notation $V_x^j = D_x V^j$ to denote the partial derivative of V^j with respect to x . Finally, notice that the terms in the last row, associated with the laws of motion for the household's employment status and the aggregate state of the economy, are outside the max operator because they are taken as given by the household.

This HJB equation holds everywhere in the interior of the household state space, i.e. for all $a > \underline{a}$ and $k > 0$. To characterize household behavior along the boundary of the state space, the HJB equation must be paired with a set of *state constraint boundary conditions*. For details, see e.g. Achdou *et al.* (2015). Intuitively, these constraints never bind in the interior of the state space. In continuous time, therefore, households that are at least an “ ϵ ” away from the constraints will never hit them in the next, infinitesimally small time step.

The household's optimization problem gives rise to the *policy functions* $c^j(a, k, \Gamma)$ and $\iota^j(a, k, \Gamma)$, which characterize household behavior at every possible point of the state space. I will characterize these policy functions next. In the following, I will frequently make use of the shorthand notation

$$\begin{aligned}
s^j &= (r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K} \right) k + e^j - q\iota^j - \psi(\iota^j, k) - c^j \\
m^j &= (\zeta - \delta)k + \iota^j,
\end{aligned}$$

where s^j and m^j respectively correspond to the drift, or expected rate of change, in households'

liquid and illiquid asset positions.

For later reference, I want to note that the household HJB equation can equivalently be written in terms of *generators* (see my discussion in the introduction to Appendix A.1). Denoting $\mathcal{A}^j = \mathcal{A}_{(a,k,z^j)}$, we have

$$(\rho + \zeta)V^j = \max_{c^j, i^j} u(c^j, H) + \mathcal{A}^j V + \mathcal{A}_\Gamma V,$$

where

$$\mathcal{A}^j f^j(a, k, \Gamma) = s^j f_a^j + m^j f_k^j + \lambda^j (f^{-j} - f^j)$$

$$\mathcal{A}_\Gamma f^j(a, k, \Gamma) = \mu_\Gamma f_\Gamma^j + \frac{1}{2} \sigma_\Gamma^T f_{\Gamma\Gamma}^j \sigma_\Gamma.$$

I will make use of this representation below.³

Optimality conditions. The optimality condition for consumption is simply given by

$$u_c(c^j, H(\Gamma)) = V_a^j(a, k, \Gamma).$$

Given the value function, we can invert this equation to solve for the policy function $c^j = c^j(a, k, \Gamma)$. Intuitively, households equate the marginal value of consuming with the marginal value of saving for the future, i.e. the marginal value of liquid wealth V_a^j .

The optimality condition for the household's investment decision is given by

$$V_k^j(a, k, \Gamma) = \left(q + \psi_{i^j}(i^j, k) \right) V_a^j(a, k, \Gamma).$$

When the adjustment cost function ψ is non-differentiable, as is the case with a kinked functional form for example, this FOC must be modified accordingly. Intuitively, households equate the marginal values of liquid and illiquid assets, accounting for the marginal rate of transformation.

³Separating the generators as I do here is feasible because there is no interaction between the diffusion coefficients of the idiosyncratic state variables and the aggregate state variables. Indeed, the idiosyncratic state variables do not load on the aggregate risk factor at all in the baseline model. If there was an interaction, the appropriate generator associated with the vector of household state variables, $\mathcal{A}_{(a,k,z,\Gamma)}$, could not be decomposed into an *idiosyncratic* and an *aggregate* component as I do here.

Euler equation for diffusion processes

While the FOCs derived in the previous subsection fully characterize optimal household behavior, it is instructive to derive a set of Euler equations. Doing so in continuous time is less straightforward than in discrete time. For illustration, I will start with the case where household earnings follow a standard diffusion process without jumps, given by

$$dz = \mu_z dt + \sigma_z dW,$$

where W is a standard Brownian motion that is uncorrelated across households. The drift and diffusion coefficients, μ_z and σ_z , are arbitrary functions on the household state space. Finally, when the earnings state variable z is continuous rather than discrete, the HJB equation that characterizes the household problem is no longer a *system* of equations. Similarly, I will drop j superscripts and denote the policy functions as $c(a, k, z, \Gamma)$ and $\iota(a, k, z, \Gamma)$ since z is now continuous.

Assumption. Crucially, I assume for now that $\mathbb{E}[dWdB] = 0$ so that idiosyncratic risk and aggregate risk are uncorrelated. I will relax this assumption in Appendix Section XX.

Plugging the policy functions back into the HJB equation (with max operator), we get a characterization of households' *indirect* value function (without max operator since optimization already took place). This HJB equation for the diffusion-only case is given by

$$(\rho + \zeta)V = u(c, H) + mV_k + sV_a + \mu_z V_z + \frac{\sigma_z^2}{2} V_{zz} + \mu_\Gamma V_\Gamma + \frac{1}{2} \sigma_\Gamma^T V_{\Gamma\Gamma} \sigma_\Gamma.$$

We can now derive what is known as the *HJB envelope condition* by differentiating with respect to a (and using the Envelope Theorem). This yields

$$(\rho - r)V_a = mV_{ka} + sV_{aa} + \mu_z V_{za} + \frac{\sigma_z^2}{2} V_{zza} + \mu_\Gamma V_{\Gamma a} + \frac{1}{2} \sigma_\Gamma^T V_{\Gamma\Gamma a} \sigma_\Gamma.$$

Next, notice that since the state variables, (a, k, z, Γ) , follow time-homogeneous Ito diffusion processes, so does any function that takes (a, k, z, Γ) as its only arguments. Thus, applying Ito's

lemma to $V_a = V_a(a, k, z, \Gamma)$ yields

$$dV_a = sV_{aa}dt + mV_{ak}dt + \mu_z V_{az}dt + \sigma_z V_{az}dW + \frac{\sigma_z^2}{2} V_{azz}dt + \mu_\Gamma V_{a\Gamma}dt + \sigma_\Gamma V_{a\Gamma}dB + \frac{1}{2}\sigma_\Gamma^T V_{a\Gamma\Gamma}\sigma_\Gamma dt.$$

Putting this equation together with the envelope condition yields

$$dV_a = (\rho - r)V_a dt + \sigma_z V_{az}dW + \sigma_\Gamma V_{a\Gamma}dB.$$

Using the household optimality condition for consumption, we have $du_c = dV_a$ and, in particular,

$$V_{az} = u_{cc}c_z$$

$$V_{a\Gamma} = u_{cc}c_\Gamma.$$

Thus, we arrive at the first of several characterizations of a continuous-time Euler equation for household consumption.

Lemma 13. *A continuous-time Euler equation for households' marginal utility of consumption is given by*

$$\frac{du_c}{u_c} = (\rho - r)dt + \sigma_z \frac{u_{cc}}{u_c} c_z dW + \sigma_\Gamma \frac{u_{cc}}{u_c} c_\Gamma dB.$$

Without taking a stance on the functional form $u(c, h)$, we cannot make much progress. Assuming CRRA preferences, however, with relative risk aversion coefficient γ implies

$$\frac{u_{cc}}{u_c} = -\frac{\gamma}{c}.$$

Similarly, notice that c is simply a function of the stochastic process u_c whose law of motion I derived above. We have $c = f(u_c) = u_c^{-\frac{1}{\gamma}}$, and so

$$\begin{aligned} dc &= -\frac{1}{\gamma}c^{1+\gamma} \left(du_c \right) + \frac{1}{2} \left(\frac{1}{\gamma} \frac{1+\gamma}{\gamma} c^{1+2\gamma} \right) (du_c)^2 \\ &= \frac{r-\rho}{\gamma} c dt + \frac{1}{\gamma} \gamma c_z \sigma_z dW + c_\Gamma \sigma_\Gamma dB + \frac{1}{2} (1+\gamma) c^{-1} \left((c_z \sigma_z)^2 + (c_\Gamma \sigma_\Gamma)^2 \right). \end{aligned}$$

I summarize in the next Lemma.

Lemma 14. *Assuming CRRA preferences, a continuous-time Euler equation for household consumption is given by*

$$\frac{dc}{c} = \underbrace{\frac{r-\rho}{\gamma} dt}_{\text{Consumption Smoothing}} + \underbrace{\frac{1+\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 dt}_{\text{Precautionary Motive: Earnings Risk}} + \underbrace{\frac{1+\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 dt}_{\text{Precautionary Motive: Aggregate Risk}} + \frac{c_z \sigma_z}{c} dW + \frac{c_\Gamma \sigma_\Gamma}{c} dB$$

This Lemma characterizes a consumption Euler equation in *differential* form. From here, it is possible to derive an exact analog to the discrete-time Euler equation by integrating up over a given time horizon.

Let $Z = \log(c) = f(c)$. Then,

$$\begin{aligned} dZ &= f'(c)dc + \frac{1}{2}f''(c)(dc)^2 = \frac{dc}{c} - \frac{1}{2}\left(\frac{dc}{c}\right)^2 \\ &= \frac{r-\rho}{\gamma} dt + \frac{\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 dt + \frac{\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 dt + \frac{c_z \sigma_z}{c} dW + \frac{c_\Gamma \sigma_\Gamma}{c} dB \end{aligned}$$

Then, writing out the stochastic differential equation in integral form, we have

$$\begin{aligned} Z_t - Z_0 &= \int_0^t \left[\frac{r-\rho}{\gamma} + \frac{\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 + \frac{\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 \right] dt + \int_0^t \frac{c_z \sigma_z}{c} dW + \int_0^t \frac{c_\Gamma \sigma_\Gamma}{c} dB \\ \log\left(\frac{c_t}{c_0}\right) &= \int_0^t \left[\frac{r-\rho}{\gamma} + \frac{\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 + \frac{\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 \right] dt + \int_0^t \frac{c_z \sigma_z}{c} dW + \int_0^t \frac{c_\Gamma \sigma_\Gamma}{c} dB \\ c_t &= c_0 \exp \left\{ \int_0^t \left[\frac{r-\rho}{\gamma} + \frac{\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 + \frac{\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 \right] dt + \int_0^t \frac{c_z \sigma_z}{c} dW + \int_0^t \frac{c_\Gamma \sigma_\Gamma}{c} dB \right\}. \end{aligned}$$

Lemma 15. *In the limit of small time steps, we have*

$$\mathbb{E}_0 [c_{t+dt}] = \exp \left\{ \frac{r-\rho}{\gamma} dt + \frac{1+\gamma}{2} \left(\frac{c_z \sigma_z}{c}\right)^2 dt + \frac{1+\gamma}{2} \left(\frac{c_\Gamma \sigma_\Gamma}{c}\right)^2 dt \right\} c_t,$$

where each of the terms on the RHS is evaluated at time t .

A.1.2 Nexus between the HJB and KF Equations

There is a deep link between the Hamilton-Jacobi-Bellman (HJB) equation that characterizes household behavior and the Kolmogorov forward (KF) equation that describes the evolution

of the cross-sectional household distribution. (For previous discussions of this link see, for example, Lasry and Lions (2007) or Achdou *et al.* (2015).)

As in the main text, I denote the cross-sectional household distribution by $g(a, k, z, \Gamma)$. The *marginal* distributions for employed and unemployed households are given by $g^j(a, k, \Gamma)$. It is well known that the law of motion of the distribution g is characterized by a KF equation, which is a particular kind of partial differential equation (PDE).

In this subsection, I discuss the nexus between the HJB and KF equations. It is intuitively quite easy to see why there would be a deep link between these equations. The HJB equation describes the behavior of households across the state space at time t . And it is through their behavior at time t , of course, that households transition to new positions in the state space at time $t + dt$. In other words, the household's decision to save or dissave at time t will determine its liquid asset position at time $t + dt$. In this sense, characterizing how the cross-sectional distribution evolves is simply a matter of aggregating up households' decisions and, thereby, the transition flows of households in the state space.

Mathematically, this relationship turns out to be characterized by *transposition*. The KF equation is given by

$$\frac{dg_t^j(a, k)}{dt} = -\partial_a \left[s^j(a, k, \Gamma_t) g_t^j(a, k) \right] - \partial_k \left[m^j(a, k, \Gamma_t) g_t^j(a, k) \right] - \lambda^j g_t^j(a, k) + \lambda^{-j} g_t^j(a, k).$$

In the spirit of my earlier discussion of *generators*, I will define the composite operator that represents this partial differential equation as

$$\mathcal{A}_j^* f^j(a, k) = -\partial_a \left[s^j(a, k, \Gamma_t) f^j(a, k) \right] - \partial_k \left[m^j(a, k, \Gamma_t) f^j(a, k) \right] - \lambda^j f^j(a, k) + \lambda^{-j} f^j(a, k),$$

so that

$$\frac{dg_t^j(a, k)}{dt} = \mathcal{A}_j^* g_t^j(a, k).$$

The following Lemma demonstrates the deep link between the KF and HJB equations.

Lemma 16. (*HJB-KF Nexus*) *The operator \mathcal{A}_j^* which defines the KF equation of this economy is the adjoint of the idiosyncratic generator \mathcal{A}_j that defines the HJB equation.*

An adjoint is to functional operators what the transpose is to matrices. Indeed, this result is of immense practical value: After discretizing the state space (a, k, z) on a grid, functional operators like ∂_a can be represented numerically by matrices. Then, finding \mathcal{A}_j^* is as simple as transposing the matrix \mathcal{A}_j that must be constructed to solve for the household value function.

A.1.3 Inflation Risk

Consumer prices are locally deterministic in my model due to price stickiness in the form of Rotemberg adjustment costs. Therefore, the baseline model features no *inflation risk*: CPI inflation at time t is t -adapted, in the sense that households have perfect foresight over the realization P_{t+dt} at time t .

This simple model feature has far-reaching implications. In particular, the evolution of the household's liquid asset position follows a non-stochastic differential equation. Since the illiquid asset position evolves similarly non-stochastically and employment status follows a Poisson rather than a diffusion process, the model's KF equation becomes a standard partial differential equation rather than a stochastic PDE. Indeed, in a model with inflation risk, the process a_t follows a stochastic differential equation and the associated KF equation becomes a stochastic PDE. This would be the case, for example, if we assume wage stickiness but allow for fully flexible consumer prices.

Recent work in the burgeoning HANK literature has become increasingly fond of assuming sticky wages instead of sticky prices (see e.g. Auclert *et al.* (2018)). In the presence of aggregate risk, such an assumption may considerably complicate the model of household behavior and the associated law of motion of the cross-sectional household distribution.

Assume that consumer prices follow a diffusion process with

$$\frac{dP}{P} = \mu_\pi dt + \sigma_\pi dB,$$

where $\sigma_\pi \neq 0$ represents inflation risk. (In the baseline model of this paper, $\sigma_\pi = 0$ of course and I denote $\pi = \mu_\pi$.) Following the steps presented in Appendix A.1.1, the associated evolution

equation for the household's liquid asset position becomes

$$da = \left[(r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K} \right) k + e - q\iota - \psi(\iota, k) - c \right] dt - a\sigma_\pi dB.$$

where I now define the real interest rate as $r = i - \mu_\pi$. The diffusion coefficient $a\sigma_\pi$ captures the household's exposure to inflation risk. It enters negatively since higher realized inflation implies a depreciation of real liquid wealth.

In the presence of inflation risk, the Kolmogorov forward equation for the evolution of the household cross-sectional distribution becomes a *stochastic* partial differential equation.

Lemma 17. (*KF with inflation risk*) *In the presence of inflation risk, the cross-sectional household distribution evolves according to*

$$\begin{aligned} \frac{dg_t^j(a, k)}{dt} = & -\partial_a \left[s^j(a, k, \Gamma_t) g_t^j(a, k) \right] - \partial_k \left[m^j(a, k, \Gamma_t) g_t^j(a, k) \right] \\ & - \lambda^j g_t^j(a, k) + \lambda^{-j} g_t^j(a, k) + \partial_a \left[a\sigma_\pi g_t^j(a, k) \right] dB. \end{aligned}$$

This equation is a special case of the general KF equation I derive in Appendix A.2. The proof can also be found there.

Inflation risk also has considerable implications for household behavior, of course. The HJB equation for household behavior becomes

$$\begin{aligned} (\rho + \zeta)V^j = & \max_{c^j, \iota^j} \left\{ u(c^j, H) + V_a^j \left[(r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K} \right) k + e^j - q\iota^j - \psi(\iota^j, k) - c^j \right] \right. \\ & \left. + \frac{1}{2} V_{aa}^j (a\sigma_\pi)^2 + V_{a\Gamma}^j (a\sigma_\pi \sigma_\Gamma) + V_k^j \left[(\zeta - \delta)k + \iota^j \right] \right\} \\ & + \lambda^j \left[V^{-j} - V^j \right] + \mu_\Gamma V_\Gamma + \frac{1}{2} \sigma_\Gamma^T V_{\Gamma\Gamma} \sigma_\Gamma. \end{aligned}$$

Two new terms emerge as a consequence of inflation risk. The first, proportional to V_{aa} , captures the uncertainty households now face over the evolution of real liquid wealth. The second, proportional to $V_{a\Gamma}$, emerges because inflation, and therefore the household's liquid wealth, covaries with the aggregate state of the economy, Γ , as both load on the aggregate risk factor dB .

A.1.4 Death Process and Blanchard (1985) Annuity Markets

How do we aggregate in this economy with death rates? I follow Blanchard (1985). Recall that the household budget constraints are given by

$$\begin{aligned}\dot{a} &= (r + \zeta)a + \left(r^k + \frac{\Pi^Q}{K}\right)k + e - q\iota - \psi(t, k) - c \equiv s \\ \dot{k} &= (\zeta - \delta)k + \iota \equiv m,\end{aligned}$$

where, as before, I suppress that the policy functions c, e, ι, s and m take (a, k, z, Γ) as their arguments.

To aggregate in this economy, we aggregate over all household cohorts that are still alive. A cohort born at time 0 has remaining size at time t given by $\zeta e^{-\zeta t}$. This is by definition. Similarly, the size of the population at time t is given by

$$\underbrace{e^{-\zeta t}}_{\text{Surviving } t=0 \text{ cohort}} + \underbrace{\int_0^t \zeta e^{-\zeta(t-s)} ds}_{\text{Surviving subsequent cohorts}} = e^{-\zeta t} + \left[e^{\zeta(s-t)} \right]_0^t = e^{-\zeta t} + e^{\zeta(t-t)} - e^{-\zeta t} = 1.$$

It is important to realize here that the remaining size at time t of a cohort born at time 0 with *initial mass* μ is given by $\mu e^{-\zeta t}$.

I can now address the important question of aggregation. Denote by $c(s, t, a, k, z)$ the consumption at time t of a household cohort born at time s . If all households of a given cohort consumed the exact same amount, then we would have

$$C_t = c(0, t) e^{-\zeta t} + \int_0^t c(s, t) \zeta e^{\zeta(s-t)} ds$$

just like in the original Blanchard (1985) paper. But now we also have within-cohort heterogeneity. Therefore, the correct aggregation is given by

$$C_t = \int c(0, t, a, k, z) e^{-\zeta t} g(0, t, a, k, z) d(a, k, z) + \int \int_0^t c(s, t, a, k, z) \zeta e^{\zeta(s-t)} ds g(s, t, a, k, z) d(a, k, z).$$

Now, of course, I want to rewrite everything exploiting the underlying stationarity. In particular, we know that conditional on (a, k, z) , households of different cohorts behave identically. And while $g(0, t)$ is very different from $g(s, t)$, we don't have to worry about this difference because,

again, households are identical conditional on the state variables. So for consumption, it's as simple as always with

$$C_t = \int c_t(a, k, z) g_t(a, k, z) d(a, k, z).$$

Now what about aggregate savings and aggregate capital accumulation? Define S_t to be aggregate liquid savings in this economy. It must be that $S_t = 0$ at all t because of the bond market clearing condition. Define A_t to be total liquid (bond) wealth at time t . We have

$$\begin{aligned} A_t &\equiv \int a \left[e^{-\zeta t} g(0, t) \right] d(a, k, z) + \int a \left[\int_0^t \zeta e^{\zeta(s-t)} g(s, t) ds \right] d(a, k, z) \\ &= \int a \left[e^{-\zeta t} g(0, t) + \int_0^t \zeta e^{\zeta(s-t)} g(s, t) ds \right] d(a, k, z). \end{aligned}$$

Differentiating with respect to t , we have

$$\begin{aligned} S_t \equiv \dot{A}_t &= \int a \left[-\zeta e^{-\zeta t} g(0, t) + e^{-\zeta t} \frac{d}{dt} g(0, t) + \zeta g(t, t) + \int_0^t \zeta \frac{\partial}{\partial t} \left(e^{\zeta(s-t)} g(s, t) \right) ds \right] d(a, k, z) \\ &= \int a \left[-\zeta e^{-\zeta t} g(0, t) + e^{-\zeta t} \frac{d}{dt} g(0, t) + \theta g(t, t) \right. \\ &\quad \left. + \int_0^t \zeta \left(-\theta e^{\zeta(s-t)} g(s, t) + e^{\zeta(s-t)} \frac{d}{dt} g(s, t) \right) ds \right] d(a, k, z). \end{aligned}$$

Splitting up these terms here, we have

$$\begin{aligned} S_t &= \zeta \int a g(t, t) - \zeta \int a \left[e^{-\zeta t} g(0, t) + \int_0^t \zeta e^{\zeta(s-t)} g(s, t) ds \right] \\ &\quad + \int a \left[e^{-\zeta t} \frac{d}{dt} g(0, t) + \int_0^t \zeta e^{\zeta(s-t)} \frac{d}{dt} g(s, t) ds \right]. \end{aligned}$$

A couple of observations are in order. The first term is simply the definition of the liquid wealth of newly born agents. By assumption, this is 0 since new household cohorts enter with zero wealth. But in principle it could be anything else (under a different birth assumption). The second term simplifies exactly to $-\zeta A_t$ as is immediately obvious. So what do we do with the third term? Let's swap the integrals so that

$$e^{-\zeta t} \left[\int a \frac{d}{dt} g(0, t) d(a, k, z) \right] + \int_0^t \zeta e^{\zeta(s-t)} \left[\int a \frac{d}{dt} g(s, t) d(a, k, z) \right] ds.$$

The Kolmogorov forward equation still holds for every cohort-specific density function, so that

we have

$$e^{-\zeta t} \left[\int a \left(\mathcal{A}_t^* g(0, t) \right) d(a, k, z) \right] + \int_0^t \zeta e^{\zeta(s-t)} \left[\int a \left(\mathcal{A}_t^* g(s, t) \right) d(a, k, z) \right] ds.$$

Notice that the adjoint operator \mathcal{A}_t^* is t -adapted here because the prices that show up in this operator are evaluated at time t (they don't vary across cohorts). And finally, I use the adjoint relationship to arrive at

$$e^{-\zeta t} \left[\int \left(\mathcal{A}_t a \right) g(0, t) d(a, k, z) \right] + \int_0^t \zeta e^{\zeta(s-t)} \left[\int \left(\mathcal{A}_t a \right) g(s, t) d(a, k, z) \right] ds,$$

where of course

$$\mathcal{A}a = s\partial_a a + m\partial_k a + \dots = s$$

where all terms except the first are 0. Therefore, putting it all together, we have

$$S_t = -\zeta A_t + e^{-\zeta t} \left[\int s_t(a, k, z) g(0, t) d(a, k, z) \right] + \int_0^t \zeta e^{\zeta(s-t)} \left[\int s_t(a, k, z) g(s, t) d(a, k, z) \right] ds.$$

Finally, I now assume that we are in a stationary equilibrium. The assumption one makes about the initialization of newly born household cohorts imply that there are differences across cohort-specific densities. But, crucially, all cohorts face the same prices. Therefore, household policy functions, conditional on (a, k, z) , no longer depend on time. We have

$$\begin{aligned} S &= -\zeta A + \int s(a, k, z) e^{-\zeta t} g(0, t, a, k, z) d(a, k, z) + \int s(a, k, z) \int_0^t \zeta e^{\zeta(s-t)} g(s, t, a, k, z) ds d(a, k, z) \\ &= -\zeta A + \int s(a, k, z) \underbrace{\left[e^{-\zeta t} g(0, t, a, k, z) + \int_0^t \zeta e^{\zeta(s-t)} g(s, t, a, k, z) ds \right]}_{\text{Distribution of all surviving households}} d(a, k, z) \end{aligned}$$

And what is the distribution of all surviving households? It is precisely the *stationary distribution* because that is the distribution of all households alive when in the stationary equilibrium.

Therefore, finally, I have proven that

$$\begin{aligned}
S &= -\zeta A + \int s(a, k, z)g(a, k, z)d(a, k, z) \\
&= -\zeta A + \int \left[\left(r + \zeta \right) a + \left(r^k + \frac{\Pi^Q}{K} \right) k + e - q\iota - \psi(\iota, k) - c \right] g(a, k, z)d(a, k, z) \\
&= \int \left[ra + \left(r^k + \frac{\Pi^Q}{K} \right) k + e - q\iota - \psi(\iota, k) - c \right] g(a, k, z)d(a, k, z)
\end{aligned}$$

where g is the stationary distribution.

A parallel argument proves that the aggregate drift of the illiquid capital stock is given by

$$\begin{aligned}
M &= -\zeta \int kg(a, k, z)d(a, k, z) + \int \left[(\zeta - \delta)k + \iota \right] g(a, k, z)d(a, k, z) \\
&= -\zeta K + (\zeta - \delta)K + \int \iota g(a, k, z)d(a, k, z) \\
&= I - \delta K,
\end{aligned}$$

where the last line uses the market clearing condition for capital production.

Overall, positive death rates have implications for the aggregation of *flows* in portfolio positions but not their outstanding stocks. That is, we still have

$$\begin{aligned}
A &= \int ag(a, k, z)d(a, k, z) \\
K &= \int kg(a, k, z)d(a, k, z).
\end{aligned}$$

A.1.5 Illustration of Walras' Law

It is instructive and insightful to work through Walras' law. In particular, I start with households' budget constraints and will derive the goods market clearing condition.

Aggregating households' liquid asset evolution equation, I get

$$\begin{aligned} \sum_j \int s^j g^j(a, k) d(a, k) &= (r + \zeta) \sum_j \int a g^j(a, k) d(a, k) + \left(r^k + \frac{\Pi^Q}{K} \right) \sum_j \int k g^j(a, k) d(a, k) \\ &\quad + \sum_j \int e^j g^j(a, k) d(a, k) - q \sum_j \int i^j g^j(a, k) d(a, k) \\ &\quad - \sum_j \int \psi(i^j, k) g^j(a, k) d(a, k) - \sum_j \int c^j g^j(a, k) d(a, k). \end{aligned}$$

Notice that for any variable $x^j(a, k)$, it is notationally equivalent to write

$$\sum_j \int x^j(a, k) g^j(a, k) d(a, k) = \int x(a, k, z) g(a, k, z) d(a, k, z).$$

Accounting for the death process as discussed in the previous subsection, aggregation yields

$$0 = -rB^G + \left(r^k + \frac{\Pi^Q}{K} \right) K + (1 - \tau^{\text{lab}})(1 - U)wH + \tau^{\text{lump}} + \tau^{\text{UI}}U - qI^H - \Psi - C$$

where U is the unemployment rate. Notice that $L = (1 - U)H$. Using the government budget constraint to solve out for τ^{lump} yields

$$0 = \left(r^k + \frac{\Pi^Q}{K} \right) K + wL + \Pi - qI^H - \Psi - C - G.$$

Since payoffs on the liquid asset, labor tax revenue and UI insurance represent direct transfers between the government and household sectors, they cancel out in this step.

Next, I substitute in for the profits of the capital and goods producing sectors, Π^Q and Π , respectively. This yields

$$\begin{aligned} 0 &= r^k K + \left(qI - I - \Phi \left(\frac{I}{K} \right) K \right) + wL + \left(Y - wL - r^k K \right) - qI^H - \Psi - C - G \\ &= -I - \Phi \left(\frac{I}{K} \right) K + Y - \Psi - C - G \end{aligned}$$

where the second line also uses the market clearing condition for new capital, $I = I^H$. I abuse notation slightly and denote the aggregate adjustment cost paid by the capital producer simply by Φ . Rearranging yields

$$Y = C + I + \Phi + \Psi + G.$$

A.2 Appendix for Section 1.4: Solution Method

This appendix provides additional details for as well as the proofs of the *analytical* results discussed in Section 1.4 of the main text.

A.2.1 Derivation of Stochastic KF Equation (Diffusion Processes)

The proof of Proposition 3 makes use of the Kolmogorov forward (KF) equation that characterizes the evolution of the cross-sectional household distribution. In this subsection, I prove the following auxiliary result. I adopt the notation introduced in 1.4 of the main text.

Lemma 18. *Assume that \mathbf{x}_t follows a time-homogeneous Ito diffusion process given by*

$$d\mathbf{x} = \mu_{\mathbf{x}}dt + \sigma_{\mathbf{x}}^W dW + \sigma_{\mathbf{x}}^B dB$$

where $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}^W$ and $\sigma_{\mathbf{x}}^B$ are functions on the household state space, (\mathbf{x}, Γ) . W_t is a standard Brownian motion that is uncorrelated across households (idiosyncratic risk), and B_t is a standard Brownian motion that is perfectly correlated across households (aggregate risk). I assume that $\mathbb{E}[dWdB] = 0$. The law of motion of the cross-sectional household distribution $g_t(\mathbf{x})$ is given by the stochastic partial differential (Kolmogorov forward) equation

$$dg(t, \mathbf{x}) = (\mathcal{A}^*g)dt + (\mathcal{B}^*g)dB$$

where \mathcal{A}^* is the adjoint of the generator \mathcal{A} associated with the HJB equation and given by

$$(\mathcal{A}f)(\mathbf{x}) = f_{\mathbf{x}}^T \mu_{\mathbf{x}} + \frac{1}{2}(\sigma_{\mathbf{x}}^W)^T f_{\mathbf{x}\mathbf{x}}(\sigma_{\mathbf{x}}^W) + \frac{1}{2}(\sigma_{\mathbf{x}}^B)^T f_{\mathbf{x}\mathbf{x}}(\sigma_{\mathbf{x}}^B),$$

and

$$(\mathcal{B}^*f)(\mathbf{x}) = - \sum_i \partial_{x_i} \left[\sigma_{x_i}^B f \right].$$

Proof. We may equivalently write

$$d\mathbf{x} = \mu_{\mathbf{x}}dt + \sigma_{\mathbf{x}} \begin{pmatrix} dB \\ dW \end{pmatrix},$$

where

$$\sigma_x = \begin{pmatrix} \sigma_x^B & \sigma_x^W \end{pmatrix}.$$

Consider any function $\phi(t, x)$. Then by Ito's lemma,

$$d\phi = \phi_t dt + \phi_x^T dx + \frac{1}{2}(dx)^T \phi_{xx}(dx),$$

where ϕ_x denotes the gradient of ϕ with respect to x , and ϕ_{xx} its Hessian. We may rewrite this as

$$d\phi = \phi_t dt + \phi_x^T \mu_x dt + \phi_x^T \sigma_x dZ + \frac{1}{2} \text{Trace} \left[\sigma_x^T \phi_{xx} \sigma_x \right],$$

where I denote

$$dZ = \begin{pmatrix} dB \\ dW \end{pmatrix}.$$

I now use the assumption $\mathbb{E}[dWdB] = 0$. This allows me to rewrite the above equation as

$$d\phi = \phi_t dt + \phi_x^T \mu_x dt + \phi_x^T \sigma_x^W dW + \phi_x^T \sigma_x^B dB + \frac{1}{2} (\sigma_x^W)^T \phi_{xx} (\sigma_x^W) dt + \frac{1}{2} (\sigma_x^B)^T \phi_{xx} (\sigma_x^B) dt$$

To illustrate this step, imagine x was 2-dimensional. Then the term in question is

$$\begin{aligned} \frac{1}{2} \text{Trace} \left[\sigma_x^T \phi_{xx} \sigma_x \right] &= \frac{1}{2} \text{Trace} \left[\begin{pmatrix} \sigma_{x_1} & \sigma_{x_2} \end{pmatrix} \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} \sigma_{x_1} \\ \sigma_{x_2} \end{pmatrix} \right] \\ &= \frac{1}{2} \text{Trace} \left[\begin{pmatrix} \sigma_{x_1} \phi_{11} + \sigma_{x_2} \phi_{21} & \sigma_{x_1} \phi_{12} + \sigma_{x_2} \phi_{22} \end{pmatrix} \begin{pmatrix} \sigma_{x_1} \\ \sigma_{x_2} \end{pmatrix} \right] \\ &= \frac{1}{2} \text{Trace} \left[\sigma_{x_1} \phi_{11} \sigma_{x_1} + \sigma_{x_2} \phi_{21} \sigma_{x_1} + \sigma_{x_1} \phi_{12} \sigma_{x_2} + \sigma_{x_2} \phi_{22} \sigma_{x_2} \right]. \end{aligned}$$

In integral representation, this becomes,

$$\begin{aligned} \phi(T, x_T) &= \phi(\tau, x_\tau) + \int_\tau^T \left[\phi_t + \phi_x^T \mu_x + \frac{1}{2} (\sigma_x^W)^T \phi_{xx} (\sigma_x^W) + \frac{1}{2} (\sigma_x^B)^T \phi_{xx} (\sigma_x^B) \right] dt \\ &\quad + \int_\tau^T \phi_x^T \sigma_x^B dB_t + \int_\tau^T \phi_x^T \sigma_x^W dW_t. \end{aligned}$$

Consider the initial condition $x_\tau = \bar{x}$. Since I picked $\phi(t, x)$ arbitrarily, it is without loss to

assume that $\phi(\tau, \bar{x}) = 0$ and $\phi(t, x) \rightarrow 0$ uniformly in x as $t \rightarrow T$. Then we have

$$\begin{aligned} & \int [\phi(T, x_T) - \phi(\tau, x_\tau)] g_t(x) dx \\ &= \int_\tau^T \int \left[\phi_t + \phi_x^T \mu_x + \frac{1}{2} (\sigma_x^W)^T \phi_{xx} (\sigma_x^W) + \frac{1}{2} (\sigma_x^B)^T \phi_{xx} (\sigma_x^B) \right] g_t(x) dx dt \\ & \quad + \int_\tau^T \int \phi_x^T \sigma_x^B g_t(x) dx dB_t + \int_\tau^T \int \phi_x^T \sigma_x^W g_t(x) dx dW_t, \end{aligned}$$

where, slightly abusing notation, the integral is taken over the state space of x . The key realization now is that, because dW_t is perfectly uncorrelated, we have

$$\int_\tau^T \int \phi_x^T \sigma_x^W g_t(x) dx dW_t = \mathbb{E}_t \left[\phi_x^T \sigma_x^W dW \right] = 0.$$

This follows because a law of large numbers holds with respect to idiosyncratic risk (but not with respect to aggregate risk).

I will now introduce an operator given by

$$(\mathcal{A}_t f)(x) = f_x^T \mu_x + \frac{1}{2} (\sigma_x^W)^T f_{xx} (\sigma_x^W) + \frac{1}{2} (\sigma_x^B)^T f_{xx} (\sigma_x^B).$$

Therefore,

$$0 = \int_\tau^T \int \left[\phi_t + (\mathcal{A}_t \phi)(x) \right] g_t(x) dx dt + \int_\tau^T \int \phi_x^T \sigma_x^B g_t(x) dx dB_t$$

I will now drop time subscript notation. Going forward, every subscript will denote a partial derivative. And recall, the second integral is taken over the region of support for x . Integrating by parts on the first term, I obtain

$$\int_\tau^T \int \phi_t g(t, x) dx dt = \int \left[\phi g \right]_\tau^T dx - \int_\tau^T \int \phi g_t dt dx,$$

where now $g_t = \partial_t g(t, x)$. Since I constructed ϕ so that $\phi(\tau) = \phi(T) = 0$, the first term here drops out. Integrating over \mathcal{A} by parts, and restricting attention to the class of functions ϕ with

compact support, I obtain

$$\begin{aligned}
& \int_{\tau}^T \int (\mathcal{A}_t \phi)(\mathbf{x}) g_t(\mathbf{x}) d\mathbf{x} dt \\
&= - \int_{\tau}^T \int \phi \left[\sum_i \partial_{x_i} (\mu_{x_i} g) - \frac{1}{2} \sum_{i,j} \partial_{x_i x_j} (\sigma_{x_i}^W g \sigma_{x_j}^W + \sigma_{x_i}^B g \sigma_{x_j}^B) \right] d\mathbf{x} dt \\
&= \int_{\tau}^T \int \phi (\mathcal{A}^* g) d\mathbf{x} dt,
\end{aligned}$$

where \mathcal{A}^* is the adjoint of \mathcal{A} .

What remains, therefore, is the aggregate risk term. Here, we have

$$\int_{\tau}^T \left[\int \phi_x^T \sigma_x^B g d\mathbf{x} \right] dB = \int_{\tau}^T \left[\phi_x^T \sigma_x^B g \Big|_{\text{Boundary}} - \int \phi \sum_i \partial_{x_i} (\sigma_{x_i}^B g) d\mathbf{x} \right] dB.$$

The first term on the RHS is again 0 due to our choice of ϕ . Therefore,

$$\int_{\tau}^T \left[\int \phi_x^T \sigma_x^B g d\mathbf{x} \right] dB = - \int_{\tau}^T \int \phi \sum_i \partial_{x_i} [\sigma_{x_i}^B g] d\mathbf{x} dB$$

In conclusion, I therefore obtain

$$0 = \int_{\tau}^T \int \phi \left[g_t - (\mathcal{A}^* g) \right] d\mathbf{x} dt + \int_{\tau}^T \int \phi \left[\sum_i \partial_{x_i} [\sigma_{x_i}^B g] \right] d\mathbf{x} dB.$$

Since this equation holds for arbitrary test functions $\phi(t, \mathbf{x})$, it must be that

$$0 = \int_{\tau}^T \phi \left[g_t - (\mathcal{A}^* g) \right] dt + \int_{\tau}^T \phi \left[\sum_i \partial_{x_i} [\sigma_{x_i}^B g] \right] dB \tag{A.1}$$

The analogous differential formulation is

$$dg(t, \mathbf{x}) = (\mathcal{A}^* g) dt - \sum_i \partial_{x_i} [\sigma_{x_i}^B g] dB.$$

This concludes the proof. ■

A.2.2 Proof of Proposition 3 (Diffusion Processes)

Proposition 3 applies to finite-dimensional approximations of the cross-sectional distribution that take the form

$$\hat{g}_t(\mathbf{x}) \equiv F(\boldsymbol{\alpha}_t)(\mathbf{x}) \approx g_t(\mathbf{x}),$$

for some $F(\cdot)$ and $\boldsymbol{\alpha}_t \in \mathbb{R}^N$. I conjecture and verify that, as long as \mathbf{x}_t and \mathbf{X}_t are time-homogeneous Ito diffusion processes, $\boldsymbol{\alpha}_t$ is similarly an N -dimensional Ito diffusion process, with

$$d\boldsymbol{\alpha}_t = \boldsymbol{\mu}_\alpha dt + \boldsymbol{\sigma}_\alpha dB,$$

where B_t is the standard, one-dimensional Brownian motion that represents the sole aggregate risk factor of the model.

By Ito's lemma,

$$\begin{aligned} d\hat{g}_t(\mathbf{x}) &= F_\alpha^T d\boldsymbol{\alpha}_t + \frac{1}{2} \sigma_\alpha^T F_{\alpha\alpha} \sigma_\alpha dt \\ &= F_\alpha^T (\boldsymbol{\mu}_\alpha dt + \boldsymbol{\sigma}_\alpha dB) + \frac{1}{2} \sigma_\alpha^T F_{\alpha\alpha} \sigma_\alpha dt, \end{aligned}$$

where F_α and $F_{\alpha\alpha}$ are the gradient and Hessian of F with respect to $\boldsymbol{\alpha}$, respectively, evaluated at $\boldsymbol{\alpha}_t$ and \mathbf{x} . Using the approximation $d\hat{g}_t(\mathbf{x}) \approx dg_t(\mathbf{x})$ as well as the Kolmogorov forward equation which I derived in Appendix A.2.1, we have

$$\begin{aligned} F_\alpha^T \boldsymbol{\mu}_\alpha dt + \frac{1}{2} \sigma_\alpha^T F_{\alpha\alpha} \sigma_\alpha dt + F_\alpha^T \boldsymbol{\sigma}_\alpha dB &\approx dg_t(\mathbf{x}) \\ &= (\mathcal{A}^* g) dt + (\mathcal{B}^* g) dB. \end{aligned}$$

Matching coefficients, we have

$$\begin{aligned} F_\alpha^T \boldsymbol{\sigma}_\alpha &= (\mathcal{B}^* g) \\ F_\alpha^T \boldsymbol{\mu}_\alpha + \frac{1}{2} \sigma_\alpha^T F_{\alpha\alpha} \sigma_\alpha &= (\mathcal{A}^* g). \end{aligned}$$

We can again make use of the approximation $F(\boldsymbol{\alpha}_t)(\boldsymbol{x}) \approx g_t(\boldsymbol{x})$ to finally arrive at

$$\begin{aligned} F_{\alpha}^T \sigma_{\alpha} &= (\mathcal{B}^* F) \\ F_{\alpha}^T \mu_{\alpha} + \frac{1}{2} \sigma_{\alpha}^T F_{\alpha\alpha} \sigma_{\alpha} &= (\mathcal{A}^* F) \end{aligned}$$

where I suppress all function inputs for convenience.

To conclude the proof of Proposition 3, it remains to solve for the $N \times 1$ vectors μ_{α} and σ_{α} . In terms of the underlying economics, this step represents an estimation or forecasting problem from the perspective of the economy's optimizing agents. There are many equally valid approaches to this estimation problem, each corresponding to a different choice of norm. In the main text, I present Proposition 3 for a *least squares* norm. Here, I derive and discuss several alternatives.

Least squares. We seek to minimize

$$\mathbb{E}_0 \left[\sum_{t=0}^T \left(\int \left[F(\boldsymbol{\alpha}_t)(\boldsymbol{x}) - g_t(\boldsymbol{x}) \right]^2 d\boldsymbol{x} \right)^{\frac{1}{2}} \right].$$

The derivation is analogous to that of the standard least squares estimator. In particular, the estimates that minimize the $\mathbb{L}^2(\boldsymbol{x})$ norm are given by

$$\begin{aligned} \sigma_{\alpha} &= (F_{\alpha}^T F_{\alpha})^{-1} F_{\alpha}^T (\mathcal{B}^* F) \\ \mu_{\alpha} &= (F_{\alpha}^T F_{\alpha})^{-1} F_{\alpha}^T \left[(\mathcal{A}^* F) - \frac{1}{2} \sigma_{\alpha}^T F_{\alpha\alpha} \sigma_{\alpha} \right]. \end{aligned}$$

Collocation. Let \mathcal{S} denote a set of collocation nodes denoted \boldsymbol{s} . If we want to pick $F(\boldsymbol{\alpha}_t)$ so that $F(\boldsymbol{\alpha}_t)(\boldsymbol{s}) = g_t(\boldsymbol{s})$ with *equality* at all collocation nodes \boldsymbol{s} , then we need as many $\boldsymbol{\alpha}_t$ as there are collocation nodes. I denote this number by N to remain consistent with the previous discussion, so that $\boldsymbol{\alpha}_t \in \mathbb{R}^N$.

At these collocation nodes, we thus have

$$\left(F_{\alpha}^T \mu_{\alpha} + \frac{1}{2} \sigma_{\alpha}^T F_{\alpha\alpha} \sigma_{\alpha} \right) dt + F_{\alpha}^T \sigma_{\alpha} dB = (\mathcal{A}^* g) dt + (\mathcal{B}^* g) dB$$

with equality as well. Matching the diffusion coefficients leads to

$$\underbrace{F_{\alpha}^T(\boldsymbol{\alpha}_t)(\mathbf{s})}_{N \times N} \underbrace{\sigma_{\alpha,t}}_{N \times 1} = \underbrace{(\mathcal{B}_t^* g_t)(\mathbf{s})}_{N \times 1}$$

where I explicitly account for the function arguments to highlight that this is a system of N equations *at* the N collocation nodes. And since $F(\boldsymbol{\alpha}_t)(\mathbf{s}) = g_t(\mathbf{s})$ exactly at the collocation nodes, we can invert and simply rewrite this as

$$\underbrace{\sigma_{\alpha}}_{N \times 1} = \underbrace{(F_{\alpha}^T)^{-1}}_{N \times N} \underbrace{(\mathcal{B}^* F)}_{N \times 1}$$

and similarly

$$\underbrace{\mu_{\alpha}}_{N \times 1} = \underbrace{(F_{\alpha}^T)^{-1}}_{N \times N} \left(\underbrace{(\mathcal{A}^* F)}_{N \times 1} - \frac{1}{2} \underbrace{\sigma_{\alpha}^T F_{\alpha\alpha} \sigma_{\alpha}}_{N \times 1} \right),$$

where the RHS of each equation is evaluated *at* the N collocation nodes \mathbf{s} .

Example. It is easiest to illustrate the previous arguments by assuming $F(\boldsymbol{\alpha}_t)(\mathbf{x})$ takes the form of a linear basis function expansion. That is, consider the approximation

$$T(\mathbf{x})\boldsymbol{\alpha}(t) \equiv \sum_{n=1}^N \alpha^n(t) T^n(\mathbf{x}) \approx g(t, \mathbf{x}),$$

where N is the number of basis functions, and the α^n are the basis function coefficients. Any linear basis function can be chosen for T^n , such as Chebyshev polynomials, hat functions, splines, etc.

For these basis functions, the second-order Ito's term conveniently drops out. That is, relating this example back to the general setting, we have

$$F_{\alpha} = \begin{pmatrix} T^1(\mathbf{x}) \\ \vdots \\ T^N(\mathbf{x}) \end{pmatrix}$$

and

$$F_{\alpha\alpha} = 0.$$

I start with the collocation approach first because it is particularly straightforward in this

setting. Denoting the N collocation nodes by s_1, \dots, s^N , we have the system of equations

$$\begin{aligned} \sum_{n=1}^N \alpha^n(t) T^n(s_1) &= g(t, s_1) \\ &\vdots \\ \sum_{n=1}^N \alpha^n(t) T^n(s_N) &= g(t, s_N) \end{aligned}$$

Taking the time differential yields

$$\begin{aligned} \sum_{n=1}^N d\alpha^n(t) T^n(s_1) &= dg(t, s_1) = (\mathcal{A}^*g)(s_1)dt + (\mathcal{B}^*g)(s_1)dB \\ &\vdots \\ \sum_{n=1}^N d\alpha^n(t) T^n(s_N) &= dg(t, s_N) = (\mathcal{A}^*g)(s_N)dt + (\mathcal{B}^*g)(s_N)dB \end{aligned}$$

Inverting this system, we have

$$d\alpha = T(\mathbf{s})^{-1} \left[(\mathcal{A}^*T)(\mathbf{s}) \alpha dt + (\mathcal{B}^*T)(\mathbf{s}) \alpha dB \right].$$

Matching coefficients then yields

$$\begin{aligned} \mu_\alpha &= T(\mathbf{s})^{-1} (\mathcal{A}^*T)(\mathbf{s}) \alpha \\ \sigma_\alpha &= T(\mathbf{s})^{-1} (\mathcal{B}^*T)(\mathbf{s}) \alpha. \end{aligned}$$

The second-order Ito term has dropped out as anticipated. In this form, it is particularly easy to see that the formulas for μ_α and σ_α only depend on terms that are readily available during the value function iteration step. In particular, there are three types of objects: First, the basis function matrix $T(\mathbf{x})$, which must be evaluated at \mathbf{s} , is chosen ex ante. Therefore, both $T(\mathbf{s})$ and $T(\mathbf{s})^{-1}$ can easily be computed before even starting the value function iteration. Second, the vector α is part of the aggregate state of the approximate economy and, therefore, part of the grid on which the value function is computed. Finally, we have the functional operators \mathcal{A}^* and \mathcal{B}^* . When the household's state space is discretized on a grid, these operators become *matrices*. And as I have repeatedly demonstrated throughout this paper, both matrices *only depend* on the

household's policy functions.

A.2.3 Choosing $F(\cdot)$ from a parametric family

I will now present several parametric families that work well in practice. For simplicity, I will present the functional forms when $x = x$ is one-dimensional. In most cases, the higher-dimensional generalizations are straightforward.

Example. (*Nodal basis functions*) The most basic class of basis functions that can be employed to great effect are nodal basis functions.⁴ Let

$$F^n(\alpha_t)(x) = \sum_{i=1}^n \alpha_{i,t} T_i(x),$$

where $T_i(x)$ is the (possibly asymmetric) hat function which takes on the value 1 on the grid point on which it is centered, and 0 on all other *nodes*. For off-node grid points, interpolation is used.

Example. (*Chebyshev polynomials*) The representation with Chebyshev polynomials takes on the same form as the nodal basis representation, except that $T_i(x)$ is now the i th Chebyshev polynomial.

Example. (*Radial basis functions*) Let $\alpha_{i,t} = \{\gamma_{i,t}, \mu_{i,t}, \sigma_{i,t}\}$, then

$$F^n(\alpha_t)(x) = \sum_{i=1}^n \gamma_{i,t} \exp\left(-\frac{(x - \mu_{i,t})^2}{2(\sigma_{i,t})^2}\right).$$

Example. (*Generalized beta density functions*) Let $B(\alpha, \beta)$ denote the beta function. Then the standard beta distribution PDF can be used with $\alpha_t \in \mathbb{R}^3$,

$$F(\alpha_t)(x) = \alpha_{1,t} \frac{x^{\alpha_{2,t}-1} (1-x)^{\alpha_{3,t}-1}}{B(\alpha_{2,t}, \alpha_{3,t})}.$$

Other functional forms from the five-parameter generalized beta family can be used as well. The beta family can be particularly efficient when modeling the household income distribution.

⁴This is precisely the basis used in the class of linear perturbation methods recently popularized by Reiter (2009).

A.2.4 Non-parametric algorithm

I will present the main argument for the special case where $F(\alpha_t)(\mathbf{x})$ is affine. Abusing notation slightly, the approximation mapping can be rendered as $g_t(\mathbf{x}) \approx C(\mathbf{x}) + \alpha_t F(\mathbf{x})$. Assuming we have simulated functional data for g_t at hand, the estimation problem may then be set up as

$$\min_{F(\mathbf{x}), C(\mathbf{x}), \alpha_t} \left\| g_t(\mathbf{x}) - C(\mathbf{x}) - \alpha_t F(\mathbf{x}) \right\|_{\mathbb{L}^2(t \times \mathbf{x})}.$$

Lemma 19. *The efficient affine basis function can be implemented using $C(\mathbf{x}) = \mathbb{E}_0(g_t(\mathbf{x}))$ and*

$$F(\mathbf{x}) = \left(\sum_t \alpha'_t \alpha_t \right)^{-1} \sum_t \alpha'_t [g_t(\mathbf{x}) - C(\mathbf{x})],$$

where $\alpha'_t = (F(\mathbf{x})F(\mathbf{x})')^{-1}F(\mathbf{x})(g_t(\mathbf{x}) - C(\mathbf{x}))'$.

The goal of Lemma 19 is to estimate an efficient basis function representation

$$F(\alpha_t)(\mathbf{x}) \approx g_t(\mathbf{x})$$

after simulating the model to obtain function data for $g_t(\mathbf{x})$. Lemma 3 tackles this problem by restricting attention, for clearer illustration, to the affine class of basis functions given by

$$C(\mathbf{x}) + T(\mathbf{x})\alpha_t.$$

The idea is to express the simulated data $g_t(\mathbf{x})$ as

$$g_t(\mathbf{x}) = C(\mathbf{x}) + T(\mathbf{x})\alpha_t + \epsilon_t(\mathbf{x}),$$

where ϵ_t is the residual error in the approximate distribution representation. It is important to note that $\epsilon_t(\mathbf{x})$ is itself a *functional* residual since we are considering an estimation problem on functional data.

Consider the problem

$$\min_{T(\mathbf{x}), C(\mathbf{x}), \alpha_t} \left\| g_t(\mathbf{x}) - C(\mathbf{x}) - T(\mathbf{x})\alpha_t \right\|_{\mathbb{L}^2(t \times \mathbf{x})}$$

where α_t is again a $N \times 1$ vector. The matrix of basis functions T can be thought of as a $J \times N$ matrix in the context of a discretized grid of J nodes (i.e. J grid points for household state \mathbf{x}).

Similarly, $C(\mathbf{x})$ can be thought of as a $J \times 1$ vector. The $\mathbb{L}^2(t \times \mathbf{x})$ norm under consideration could correspondingly be recast as

$$\min_{T_{nj}, C_j, \alpha_{tn}} \sum_{t,j} \left(g_{tj} - C_j - \sum_n \alpha_{tn} T_{nj} \right)^2.$$

In terms of the residual simulation error mentioned in the previous paragraph, this loss criterion is equivalent to $\sum_i \epsilon_{ti}^2 = \sum_i \epsilon_t(x_i)^2 = \epsilon_t'(\mathbf{x}) \epsilon_t(\mathbf{x})$ for a given t .

I will set the constant equal to the unconditional mean of the simulated distribution.

$$C = \mathbb{E}_T(g_t)$$

Let $\hat{g}_t = g_t - C$. Then we have

$$\min \sum_t (\hat{g}_t - T\alpha_t)' (\hat{g}_t - T\alpha_t).$$

Factoring out,

$$\begin{aligned} \min \sum_t & \left(\hat{g}_t' \hat{g}_t - \hat{g}_t' (T\alpha_t) - (T\alpha_t)' \hat{g}_t + (T\alpha_t)' (T\alpha_t) \right) \\ & \sum_t \left(\hat{g}_t' \hat{g}_t - 2(T\alpha_t)' \hat{g}_t + (T\alpha_t)' (T\alpha_t) \right) \end{aligned}$$

Differentiating with respect to α_t yields

$$\begin{aligned} 0 &= -2 \frac{\partial}{\partial \alpha_t} (T\alpha_t)' \hat{g}_t + \frac{\partial}{\partial \alpha_t} (T\alpha_t)' (T\alpha_t) \\ &= -2 \underbrace{T' \hat{g}_t}_{N \times 1} + \underbrace{(T'T + (T'T)')}_{N \times 1} \alpha_t. \end{aligned}$$

Rearranging, we have

$$(T'T + (T'T)') \alpha_t = 2T\hat{g}_t$$

Finally, note that $T'T = (T'T)'$ so that we have

$$\alpha_t = (T'T)^{-1} T' \hat{g}_t.$$

This is, of course, precisely consistent with the estimation problem for the law of motion $d\alpha_t$, which is the subject of Proposition 3.

Next, consider the optimality condition for $T(\mathbf{x})$. Using

$$\frac{\partial}{\partial \mathbf{X}} \mathbf{b}' \mathbf{X}' \mathbf{D} \mathbf{X} \mathbf{c} = \mathbf{D}' \mathbf{X} \mathbf{b} \mathbf{c}' + \mathbf{D} \mathbf{X} \mathbf{c} \mathbf{b}'.$$

we have

$$0 = \sum_t \frac{\partial}{\partial T} \left[\hat{\mathbf{g}}_t \hat{\mathbf{g}}_t' - 2(\mathbf{T} \boldsymbol{\alpha}_t)' \hat{\mathbf{g}}_t + (\mathbf{T} \boldsymbol{\alpha}_t)' (\mathbf{T} \boldsymbol{\alpha}_t) \right]$$

which becomes

$$0 = \sum_t \left[-2\hat{\mathbf{g}}_t \boldsymbol{\alpha}_t' + 2\mathbf{T}(\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t') \right],$$

or simply

$$0 = \sum_t \left[(\hat{\mathbf{g}}_t - \mathbf{T} \boldsymbol{\alpha}_t) \boldsymbol{\alpha}_t' \right].$$

Rearranging, we can solve for $T(\mathbf{x})$ as

$$T(\mathbf{x}) = \left(\sum_t \boldsymbol{\alpha}_t \boldsymbol{\alpha}_t' \right)^{-1} \sum_t \hat{\mathbf{g}}_t \boldsymbol{\alpha}_t'$$

This concludes the proof.

A.3 Appendix for Section 1.5: Data and Empirics

A.3.1 Employment Transitions in CPS Micro Data

In this subsection, I provide details and additional material for my estimation of employment transition rates using Current Population Survey (CPS) data.

Raw CPS data

The Current Population Survey (CPS) is a household survey conducted by the Bureau of Labor Statistics (BLS) at a monthly frequency since 1948. The survey features a rotating panel of households. A household is typically in the survey for four consecutive months. It is therefore possible to use a household identifier to match records across months and create a panel of household employment transitions. There is a long tradition of constructing gross employment flows from matched CPS micro data. See for example Poterba and Summers (1986), Blanchard *et al.* (1990) and Shimer (2012).

I obtain the raw CPS data from the NBER.⁵ I restrict my sample to the period 1996 through 2019 throughout the analysis. A key question I use to construct employment categories was only introduced to the CPS in 1994. Furthermore, the years 1994 and 1995 saw several changes in, for example, the definition of household identifiers. This is a well-known issue that has made it impossible to match data across several months in 1994 and 1995. I therefore start my sample in 1996. Similarly, I end the sample in December 2019, before the onset of the Covid pandemic in the U.S.

Employment status definitions

The conventional definition of unemployment used by the BLS counts those above age 16 who are currently unemployed but actively seeking employment, as a percent of the civilian labor force. This notion is also called headline unemployment or U3 unemployment.

U3 has long been a contentious definition of unemployment and is oftentimes criticized as being too narrowly defined. Indeed, it is well known that groups such as the *marginally attached*

⁵See <https://data.nber.org/data/cps-basic2/>.

exhibit transition behavior that is significantly distinct from a more narrowly defined *not in the labor force* (NILF) group and more similar to the behavior of the unemployed. Specifically, the marginally attached (and some other groups) are much more likely to transition into employment or unemployment than discouraged workers, the disabled or the retired, all of whom are lumped together in the NILF category.

My model calls for a definition of employment and unemployment that captures as large a share as possible of those households who exhibit positive employment transition rates. Since the model speaks to questions about *aggregate* consumer behavior, it is only fitting to include as large a share of the total (prime-age) population as is reasonable when calibrating its parameters.

On the other hand, there is also a strong argument to *exclude* households that are not currently in the labor force and are sufficiently unlikely to transition back. To include such households, for example the disabled or retired, it would be more consistent to define a distinct employment state, in which households are not exposed to employment transitions at all. As I want to restrict the model to two employment states, the most reasonable compromise is to define a non-employment group that includes all those households that currently face a sufficiently high probability of transitioning back into employment.

Concretely, I let employment state E correspond to the conventional definition of *employment*. On the other hand, I define the model's unemployment state U as the union of the conventionally unemployed, the marginally attached, and those employed part-time involuntarily for economic reasons. I furthermore restrict the sample to prime-age workers. Therefore, my overall definition of unemployment is more similar to the BLS' U6 measure.

I define the *marginally attached* as those households currently not in the labor force that (i) want a job, (ii) have looked for employment in the last 12 months, and (iii) would be able to take a job next week if offered.⁶ Similarly, I categorize households as unemployed part-time for economic reasons if they are currently working fewer than 35 hours per week but would like to work more if given the chance.⁷

⁶The corresponding CPS variables are PRWNTJOB, PEDWLKO and PEDWAVL.

⁷The corresponding CPS variable is PEHRRSN1.

Creating monthly panel data

To match households across their months in the survey and create an associated panel, I follow the steps in Shimer (2012).⁸ In particular, to match households I use the household identifier variables provided in the CPS, variables for race, sex and age, as well as the household's line item number and months in the survey.⁹

From gross flows to continuous-time transition rates

After taking the steps discussed thus far, one can compute the monthly *gross flows* across employment states. I now discuss how to map discrete-time flow data at a monthly frequency to continuous-time transition rates at a quarterly frequency. The following derivations are closely related to those in Shimer (2012) who argues that using continuous-time transition rates can circumvent a time aggregation bias inherent in discrete-time data.

Let j and k denote employment states, $j, k \in \{U, E\}$. The time horizon is $t \in [0, \infty)$ at a quarterly frequency. Denote by $N_t^{jk}(\tau)$ the number of households in employment state j at time t and state k at time $t + \tau$. Define

$$n_t^{jk}(\tau) = \frac{N_t^{jk}(\tau)}{N_t^{jj}(\tau) + N_t^{jk}(\tau)}$$

as the share of all those households in employment state j at time t that have transitioned into state k at time $t + \tau$. Naturally, $n_t^{jk}(0) = N_t^{jk}(0) = 0$ for $k \neq j$, and $n_t^{jj}(\tau) = 1 - n_t^{jk}(\tau)$. Finally, I denote by λ_t^j the continuous-time transition rate *out of* state j at time t . This is the same transition rate that appears in the quantitative model. In particular, λ_t^E is the job separation rate and λ_t^U the job finding rate.

Lemma 20. *The relationship between flow shares and transition rates is given by*

$$n_t^{jk}(\tau) = \lambda_t^j \frac{1 - e^{-(\lambda_t^j + \lambda_t^k)\tau}}{\lambda_t^j + \lambda_t^k}.$$

⁸My code builds on the replication code files provided by Robert Shimer, <https://sites.google.com/site/robertshimer/research/flows>. For additional details, please see Shimer (2012).

⁹Households are in the CPS for four consecutive months, then leave the survey and subsequently come back for a second rotation of four consecutive months. I use data from both rounds.

Lemma 21. *Data on discrete-time gross flows at monthly frequency (from the CPS) can be mapped into continuous-time Poisson transition rates at quarterly frequency using the formulas*

$$\lambda^U = -\frac{n^{UE}}{n^{UE} + n^{EU}} \cdot 3 \log \left(1 - n^{EU} - n^{UE} \right)$$

$$\lambda^E = -\frac{n^{EU}}{n^{UE} + n^{EU}} \cdot 3 \log \left(1 - n^{EU} - n^{UE} \right).$$

A.3.2 Business Cycle Moments

The quantitative model generates naturally asymmetric business cycles and closely matches prominent features of the business cycles in U.S. postwar history. In this section, I discuss the data I use to compute these business cycle moments.

Raw data

The primary series I consider are output, investment, consumption and hours worked. Investment is defined as non-residential fixed investment and durable goods consumption. Consumption is defined as expenditures on non-durable goods and services. Output, consumption and investment are from the NIPA accounts. For hours, I use hours of all persons in the nonfarm business sector from the BLS.

These series are expressed in real terms after deflating by the CPI index. Similarly, all variables are expressed per capita, dividing by the civilian non-institutionalized population aged 16 and over, and in logs.

I consider a sample from the first quarter of 1953 until the fourth quarter of 2019. I exclude the immediate postwar years as they featured a degree of volatility in, for example, GDP growth that is uncharacteristic for the postwar period as a whole.

Finally, I detrend all series using an HP filter, as is standard. I use a smoothing parameter of $\lambda = 1600$, which is appropriate for quarterly data.

A.3.3 Macroeconomic Uncertainty Indices

Since the seminal work by Bloom (2009), there has been a surge in interest in identifying and measuring macro uncertainty in the data. Numerous proxies and indices have been proposed.

To confront my model's predictions about uncertainty with data, I use a whole range of empirical macro uncertainty proxies.

- Jurado *et al.* (2015) provide several direct econometric estimates of macro uncertainty. They interpret uncertainty as conditional volatility in a series' unforecastable component and macro uncertainty in particular as the *common* variation in a host of aggregate time series. Using close to 300 macroeconomic and financial time series, they estimate factor-based forecasting models and compute a macroeconomic uncertainty index (as well as financial and real uncertainty indices) at 1, 3 and 12-month horizons.¹⁰
- The Chicago Board Options Exchange (CBOE)'s Volatility Index (VIX) is a commonly used proxy for expected volatility in financial markets. Using S&P 500 index options, the index computes a one-month forward looking market expectation of volatility.
- The Economic Policy Uncertainty Index I use is from Baker *et al.* (2016).¹¹ Based on newspaper coverage, this index tallies the frequency of articles in leading U.S. newspapers that “contain the following triple: ‘economic’ or ‘economy’; ‘uncertain’ or ‘uncertainty’; and one or more of ‘congress’, ‘deficit’, ‘Federal Reserve’, ‘legislation’, ‘regulation’ or ‘White House’.”
- Finally, the World Uncertainty Index for the United States is from Ahir *et al.* (2018).¹² The index is based on the frequency with which the word “uncertainty” appears in the Economist Intelligent Unit's quarterly country reports.

¹⁰The raw time series can be found on Sydney Ludvigson's website, <https://www.sydneyludvigson.com/>.

¹¹I access the data using FRED.

¹²I also access this time series using FRED.

Appendix B

Appendix to Chapter 3

B.1 Proofs

B.1.1 Proof of Proposition 4

To verify given our conjectured mechanism, we verify the mechanism is locally incentive compatible at the constrained efficient allocation by studying one shot deviations. Adopting the sequence form of the problem, we have the (WLOG) date 0 value function

$$\mathcal{W}(v_{-1}, E_{-1}\pi, \theta) = \sup_{\pi_0, \tilde{\theta}_0} E_0 \sum_{t=0}^{\infty} \beta^t \left[-v_{t-1} (\pi_t - E_{t-1} [\pi_t | v_{t-2}, \tilde{\theta}_{t-1}]) \right. \\ \left. + U_t (\pi_t, E_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t], \theta_t) \right].$$

The first order condition for π_0 is

$$0 = -v_{-1} + \frac{\partial U_0}{\partial \pi_0}$$

giving the first order condition for inflation. Next, consider a change in the report $\tilde{\theta}_0$. By the Envelope Theorem, we have

$$\frac{\partial \mathcal{W}_0}{\partial \tilde{\theta}} = -E_0 \sum_{t=0}^{\infty} \beta^{t+1} \frac{\partial v_t}{\partial \tilde{\theta}_0} (\pi_{t+1} - E_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]) \\ + E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{dU_t (\pi_t, E_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t], \theta_t)}{dE_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]} + \beta v_t \right] \frac{dE_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]}{d\tilde{\theta}_0}$$

By Law of Iterated Expectations, we obtain

$$\begin{aligned} \frac{\partial \mathcal{W}_0}{\partial \tilde{\theta}} &= -E_0 \sum_{t=0}^{\infty} \beta^{t+1} \frac{\partial v_t}{\partial \tilde{\theta}_0} E_t \left[\pi_{t+1} - E_t [\pi_{t+1} | v_{t-1}, \tilde{\theta}_t] \middle| \theta^t \right] \\ &\quad + E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{dU_t(\pi_t, E_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t], \theta_t)}{dE_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]} + \beta v_t \right] \frac{dE_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]}{d\tilde{\theta}_0} \end{aligned}$$

Now, substituting in at the constrained efficient allocation along a truthful equilibrium, we have $\frac{dU_t(\pi_t, E_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t], \theta_t)}{dE_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t]} + \beta v_t = 0$ and $E_t[\pi_{t+1} - E_t[\pi_{t+1} | v_{t-1}, \tilde{\theta}_t] \middle| \theta^t] = 0$. Therefore, we have the first order condition in the report $\tilde{\theta}$ also holding.

As a result, this mechanism implements the constrained efficient allocation.

B.1.2 Proof of Lemma 6

Suppose that we take the Bellman equation:

$$\mathcal{W}_t(\theta^t) = \int_{\underline{\theta}}^{\theta_t} \frac{\partial U_t(\theta^{t-1}, s_t)}{\partial s_t} ds_t + \beta \int_{\underline{\theta}}^{\theta_t} E_t \left[\mathcal{W}_{t+1} \frac{\partial f_t(\theta_{t+1} | s_t) / \partial s_t}{f_t(\theta_{t+1} | s_t)} \middle| s_t \right]$$

And iterate it forward once. Iterating forward once, we obtain:

$$\begin{aligned} \mathcal{W}_t(\theta^t) &= \int_{\underline{\theta}}^{\theta_t} E_t \left[\frac{\partial U_t(\theta^{t-1}, s_t)}{\partial s_t} ds_t + \frac{\partial f_t(\theta_{t+1} | s_t) / \partial s_t}{f_t(\theta_{t+1} | s_t)} \beta \times \right. \\ &\quad \left. \left[\int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial U_t(\theta^{t-1}, s_t, s_{t+1})}{\partial s_{t+1}} + E_{t+1} \mathcal{W}_{t+2} \frac{f_{t+1}(\theta_{t+2} | s_{t+1}) / \partial s_{t+1}}{f_{t+1}(\theta_{t+2} | s_{t+1})} \middle| s_{t+1} \right] \right] \end{aligned}$$

Iterating forward, suppose that we define the following recursive operator. In particular, we define:

$$\mathcal{B}_t^0(g, \theta) = \int_{\underline{\theta}}^{\theta} g ds_t$$

Note that for the function $g_t^0 = \frac{\partial U_t(\theta^{t-1})}{\partial s_t}$, we have that \mathcal{B}_t^0 is the first term in the infinite series defining \mathcal{W}_t .

And suppose we define next:

$$\mathcal{B}_t^1(g, \theta) = \int_{\underline{\theta}}^{\theta} E_t \left[\frac{\partial f_t(\theta_{t+1} | s_t) / \partial s_t}{f_t(\theta_{t+1} | s_t)} g \middle| s_t \right] ds_t$$

Consider the function $g_t^1 = \int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial U_{t+1}(\theta^{t-1}, s_t, s_{t+1})}{\partial s_{t+1}} ds_{t+1}$. Taking the function $\mathcal{B}_t^1(g_t^1, \theta_t)$ and multiplying by β , we obtain the second term in the infinite series for \mathcal{W}_t .

From here, we define a recursive operator. Consider a function g_t^s that is a date $t + s$ adapted function. We define the operator:

$$\mathcal{B}_t^2(g_t^2, \theta_t) = \mathcal{B}_t^1\left(\mathcal{B}_{t+1}^1(g_t^2, \theta_{t+1}), \theta_t\right)$$

So that we have:

$$\begin{aligned} \mathcal{B}_t^2(g_t^2, \theta_t) = & \\ \int_{\underline{\theta}}^{\theta_t} E_t \left[\frac{\partial f_t(\theta_{t+1}|s_t)/\partial s_t}{f_t(\theta_{t+1}|s_t)} \int_{\underline{\theta}}^{\theta_{t+1}} E_{t+1} \left[\frac{\partial f_{t+1}(\theta_{t+2}|s_{t+1})/\partial s_{t+1}}{f_{t+1}(\theta_{t+2}|s_{t+1})} g_t^2(s_{t+1}, \theta_{t+2}) \Big|_{s_{t+1}} \right] ds_{t+1} \Big|_{s_t} \right] ds_t \end{aligned}$$

Which, when $g_t^2(s_t, s_{t+1}, \theta_{t+2}) = \int_{\underline{\theta}}^{\theta_{t+2}} \frac{\partial U_{t+2}(\theta^{t-1}, s_t, s_{t+1}, s_{t+2})}{\partial s_{t+2}} ds_{t+2}$ and multiplied by β^2 , gives us the next term in the infinite series defining \mathcal{W}_t .

Continuously defining these recursive operators as such, and defining functions

$$g_t^s(s_t, \dots, s_{t+s-1}, \theta_{t+s}) = \int_{\underline{\theta}}^{\theta_{t+s}} \frac{\partial U_{t+s}(\theta^{t-1}, s_t, \dots, s_{t+s})}{\partial s_{t+s}},$$

we obtain the infinite series that characterizes \mathcal{W}_t .

In other words, we can construct such recursive operators. From here, we look to simplify these operators. Let us start from the operator $\mathcal{B}_t^1(g, \theta_t)$. In particular, we have:

$$\begin{aligned} \mathcal{B}_t^1(g, \theta_t) &= \int_{\underline{\theta}}^{\theta_t} E_t \left[\frac{\partial f_t(\theta_{t+1}|s_t)/\partial s_t}{f_t(\theta_{t+1}|s_t)} g(s_t, \theta_{t+1}) \Big|_{s_t} \right] ds_t \\ &= \int_{\underline{\theta}}^{\theta_t} \int_{\theta_{t+1}} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} g(s_t, \theta_{t+1}) d\theta_{t+1} ds_t \\ &= \int_{\theta_{t+1}} \left[\int_{\underline{\theta}}^{\theta_t} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} g(s_t, \theta_{t+1}) ds_t \right] d\theta_{t+1} \\ &= \int_{\theta_{t+1}} \frac{\left[\int_{\underline{\theta}}^{\theta_t} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} g(s_t, \theta_{t+1}) ds_t \right]}{f_t(\theta_{t+1}|\theta_t)} f_t(\theta_{t+1}|\theta_t) d\theta_{t+1} \\ &= E_t \left[\frac{1}{f_t(\theta_{t+1}|\theta_t)} \left[\int_{\underline{\theta}}^{\theta_t} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} g(s_t, \theta_{t+1}) ds_t \right] \Big|_{\theta_t} \right] \end{aligned}$$

In particular, as applied to the function $g_t^1 = \int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial U_{t+1}(\theta^{t-1}, s_t, s_{t+1})}{\partial s_{t+1}} ds_{t+1}$, we obtain:

$$\mathcal{B}_t^1(g, \theta_t) = E_t \left[\frac{1}{f_t(\theta_{t+1}|\theta_t)} \left[\int_{\underline{\theta}}^{\theta_t} \int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial U_{t+1}(\theta^{t-1}, s_t, s_{t+1})}{\partial s_{t+1}} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} ds_{t+1} ds_t \right] \Big|_{\theta_t} \right]$$

Which is of the form in the Lemma.

Now, let us consider the second operator. We have:

$$\mathcal{B}_t^2(g, \theta_t) = \mathcal{B}_t^1\left(\mathcal{B}_{t+1}^1(g, \theta_{t+1}), \theta_t\right)$$

Recall that the simplified operator above expresses:

$$\mathcal{B}_t^1(g, \theta_t) = E_t \left[\frac{1}{f_t(\theta_{t+1}|\theta_t)} \left[\int_{\underline{\theta}}^{\theta_t} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} g(s_t, \theta_{t+1}) ds_t \right] \middle| \theta_t \right]$$

In other words, we have along history (θ^{t-1}, s_t) :

$$\mathcal{B}_{t+1}^1(g, \theta_{t+1}) = E_{t+1} \left[\frac{1}{f_{t+1}(\theta_{t+2}|\theta_{t+1})} \left[\int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial f_{t+1}(\theta_{t+2}|s_{t+1})}{\partial s_{t+1}} g(s_t, s_{t+1}, \theta_{t+2}) ds_{t+1} \right] \middle| \theta_{t+1} \right]$$

And applying this into the operator defining \mathcal{B}_t^2 , we obtain:

$$\mathcal{B}_t^2(g, \theta_t) = E_t \left[\frac{1}{f_t(\theta_{t+1}|\theta_t) f_{t+1}(\theta_{t+2}|\theta_{t+1})} \left[\int_{\underline{\theta}}^{\theta_t} \int_{\underline{\theta}}^{\theta_{t+1}} \frac{\partial f_t(\theta_{t+1}|s_t)}{\partial s_t} \frac{\partial f_{t+1}(\theta_{t+2}|s_{t+1})}{\partial s_{t+1}} g(s_t, s_{t+1}, \theta_{t+2}) ds_{t+1} ds_t \right] \middle| \theta_t \right]$$

And substituting in $g_t^2 = \int_{\underline{\theta}}^{\theta_{t+2}} \frac{\partial U_{t+2}(\theta^{t-1}, s_t, s_{t+1}, s_{t+2})}{\partial s_{t+2}} ds_{t+2}$, we get the next expression from the

Lemma. From here, the result follows from repeated iteration.

B.1.3 Proof of Lemma 7

For any allocation rule, T_t provides the implementation. Recall that the government's welfare is given by:

$$\max E_{-1} \left[\sum_{t=0}^{\infty} \beta^t U_t - \kappa T_t \right],$$

Recall that bank welfare is given by:

$$\mathcal{W}_0 = E_0 \sum_{t=0}^{\infty} [\beta^t U_t + T_t]$$

In other words, we always have:

$$-E_0 \sum_{t=0}^{\infty} T_t = E_0 \sum_{t=0}^{\infty} \beta^t U_t - \mathcal{W}_0$$

Substituting in above, by Law of Iterated Expectations we obtain the planning problem:

$$\max E_{-1} \left[-\kappa \mathcal{W}_0 + \sum_{t=0}^{\infty} \beta^t (1 + \kappa) U_t \right],$$

and where lastly, we use Lemma 4 substitute in for \mathcal{W}_0 to obtain the result.

B.1.4 Proof of Proposition 8

Recall that our objective function for the second-best optimization problem was given by:

$$\max \int_{\theta_0} \left[\sum_{t=0}^{\infty} \beta^t \left[-\frac{\kappa}{1 + \kappa} \mathcal{B}_0^s (g_0^t, \theta_0) + U_t (\pi_t, \pi_{t+1}, \theta_t, \theta_t) \right] \right] dF_0(\theta_0)$$

Note that given the optimal mechanism implements truthful reporting, we may substitute in $\tilde{\theta}_t = \theta_t$.

Recall further the simplified form of the operators:

$$\mathcal{B}_t^s = E_t \left[\prod_{k=0}^{s-1} \frac{1}{f_{t+k}(\theta_{t+k+1} | \theta_{t+k})} \int_{s_t \leq \theta_t, \dots, s_{t+s} \leq \theta_{t+s}} \frac{\partial U_{t+s}(\theta^{t-1}, s_t, \dots, s_{t+s})}{\partial s_{t+s}} \prod_{k=0}^{s-1} \frac{\partial f_{t+k}(\theta_{t+k+1} | s_{t+k})}{\partial s_{t+k}} ds_{t+s} \dots ds_t \middle| \theta_t \right]$$

Now, denote the *realized value* of the operator \mathcal{B}_0^t by:

$$B_0^t(\theta^t) = \prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1} | \theta_k)} \int_{s_0 \leq \theta_0, \dots, s_t \leq \theta_t} \frac{\partial U_t(s_0, \dots, s_t)}{\partial s_t} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1} | s_k)}{\partial s_k} ds_t \dots ds_0$$

So that $B_0^t(\theta^t)$ is a random variable derived from the history θ^t of shocks. Given the definition of this random variable, denote E_{-1} to be the beginning-of-period-0 expectation, not conditional on the information θ_0 . From here, we can rewrite the objective function of the government as:

$$\max E_{-1} \left[\sum_{t=0}^{\infty} \beta^t \left[-\frac{\kappa}{1 + \kappa} B_0^t(\pi_t, \pi_{t+1}, \theta_t | \theta^{t-1}) + (1 + \kappa) U_t(\pi_t, \pi_{t+1}, \theta_t | \theta_t) \right] \right]$$

From here, consider the optimal choice of inflation $\pi_t(z^t)$, for a realized history $\theta^t = z^t$ of shocks. Note that the solution can be written in the form (for $t \geq 1$):

$$\frac{\partial U_{t-1}}{\partial \pi_t(z^t)} f(z^{t-1}) + \beta \frac{\partial U_t}{\partial \pi_t(z^t)} f(z^t) = \frac{\kappa}{1 + \kappa} E_{-1} \sum_{s=t-1}^t \beta^{s-(t-1)} \frac{d}{d\pi_t(z^t)} B_0^s(\pi_s, \pi_{s+1}, \theta_s | \theta^s)$$

So that all that remains is to characterize the derivatives of B_0^s with respect to $\pi_t(z^t)$. When

$s = t$, we have:

$$\frac{d}{dz^t} B_0^t(\theta^t) = \frac{d}{\pi_t(z^t)} \left[\prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \int_{s_0 \leq \theta_0, \dots, s_t \leq \theta_t} \frac{\partial U_t(s_0, \dots, s_t)}{\partial s_t} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1}|s_k)}{\partial s_k} ds_t \dots ds_0 \right]$$

Note that $\pi_t(z^t)$ appears in $\frac{\partial U_t(s_0, \dots, s_t)}{\partial s_t}$ only along the path given by $s_0 = z_0, s_1 = z_1, \dots, s_t = z_t$. Essentially then, this derivative at a single point $\pi_t(z^t)$ comes down to extracting the derivative along that path under the integral. The derivative along that path is then given by:

$$\frac{d}{dz^t} B_0^t(\theta^t) = \mathbf{1}_{z_0 \leq \theta_0, \dots, z_t \leq \theta_t} \prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \frac{\partial^2 U_t}{\partial z_t \partial \pi_t(z^t)} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k}$$

Note the subtlety that the θ 's are preserved, as the realization of the random history, whereas the s 's are replaced by z 's, as the path under the integrals that leads to the history z^t under the integrals. It is worth remembering then, when we substitute into the expectation, that θ_t is a random variable, and z^t is (fixed) the history being differentiated along, and so is not a random variable.

Note that by exactly the same logic, we obtain $\forall t \geq 2$

$$\frac{d}{dz^t} B_0^{t-1}(\theta^{t-1}) = \mathbf{1}_{z_0 \leq \theta_0, \dots, z_{t-1} \leq \theta_{t-1}} \prod_{k=0}^{t-2} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \frac{\partial^2 U_{t-1}}{\partial z_{t-1} \partial \pi_t(z^t)} \prod_{k=0}^{t-2} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k}$$

As a result, the right-hand side of the first-order condition becomes $\forall t \geq 2$

$$\begin{aligned} \frac{1+\kappa}{\kappa} \text{RHS} &= E_{-1} \sum_{s=t-1}^t \frac{d}{d\pi_t(z^t)} B_0^s(\pi_s, \pi_{s+1}, \theta_s | \theta^s) \\ &= E_{-1} \left[\mathbf{1}_{z_0 \leq \theta_0, \dots, z_{t-1} \leq \theta_{t-1}} \prod_{k=0}^{t-2} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \frac{\partial^2 U_{t-1}}{\partial z_{t-1} \partial \pi_t(z^t)} \prod_{k=0}^{t-2} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k} \right] \\ &\quad + \beta E_{-1} \left[\mathbf{1}_{z_0 \leq \theta_0, \dots, z_t \leq \theta_t} \prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \frac{\partial^2 U_t}{\partial z_t \partial \pi_t(z^t)} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k} \right] \\ &= \frac{\partial^2 U_{t-1}}{\partial z_{t-1} \partial \pi_t(z^t)} E_{-1} \left[\mathbf{1}_{z_0 \leq \theta_0, \dots, z_{t-1} \leq \theta_{t-1}} \prod_{k=0}^{t-2} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k} \right] \\ &\quad + \frac{\partial^2 U_t}{\partial z_t \partial \pi_t(z^t)} \beta E_{-1} \left[\mathbf{1}_{z_0 \leq \theta_0, \dots, z_t \leq \theta_t} \prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1}|\theta_k)} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1}|z_k)}{\partial z_k} \right] \end{aligned}$$

Where here, we applied the fact that we have chosen a specific history z^t , so that the cross-partial above are *not* random variables, but rather are specific realizations of those random

variables. By contrast, the part inside the expectation corresponds to histories which contain these specific histories, and so are random variables.

Now, consider these two expectations. Now, we define $\Omega_t(z^t)$ by:

$$\begin{aligned}
\Omega_t(z^t) &\equiv E_{-1} \left[\mathbf{1}_{z_0 \leq \theta_0, \dots, z_t \leq \theta_t} \prod_{k=0}^{t-1} \frac{1}{f_k(\theta_{k+1} | \theta_k)} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1} | z_k)}{\partial z_k} \right] \\
&= \int_{z_t}^{\bar{\theta}} \int_{z_{t-1}}^{\bar{\theta}} \dots \int_{z_0}^{\bar{\theta}} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1} | z_k)}{\partial z_k} f(\theta_0) d\theta_t \dots d\theta_0 \\
&= \int_{z_t}^{\bar{\theta}} \frac{\partial f_k(\theta_t | z_{t-1})}{\partial z_k} \left[\int_{z_{t-1}}^{\bar{\theta}} \dots \int_{z_0}^{\bar{\theta}} \prod_{k=0}^{t-1} \frac{\partial f_k(\theta_{k+1} | z_k)}{\partial z_k} f(\theta_0) d\theta_{t-1} \dots d\theta_0 \right] d\theta_t \\
&= \int_{z_t}^{\bar{\theta}} \frac{\partial f_k(\theta_t | z_{t-1})}{\partial z_{t-1}} \Omega_{t-1}(z^{t-1}) d\theta_t \\
&= \Omega_{t-1}(z^{t-1}) \int_{z_t}^{\bar{\theta}} \frac{\partial f_k(\theta_t | z_{t-1})}{\partial z_{t-1}} d\theta_t
\end{aligned}$$

Which is well-defined for all $t \geq 1$. However, it requires an initial condition $\Omega_0(z^0)$. It is helpful to define this initial condition in the date 1 FOC. Note that at date 1, we have:

$$\mathcal{B}_0^{t-1}(\theta^{t-1}) = \mathcal{B}_0^0(\theta^0) = \int_{\underline{\theta}}^{\theta_0} \frac{\partial U_0}{\partial s_0} ds_0$$

So that we have $\frac{d}{d\pi_t(z^t)} \mathcal{B}_0^{t-1}(\theta^{t-1}) = \mathbf{1}_{z_0 \leq \theta_0} \frac{\partial U_0}{\partial \pi_1(z^1)}$. In particular then, the expectation is simply:

$$E_{-1} [\mathbf{1}_{z_0 \leq \theta_0}] = \int_{z_0}^{\bar{\theta}} f(\theta_0) d\theta_0 = 1 - F(z_0)$$

So that we have initial condition $\Omega_0(z^0) = 1 - F(z_0)$.

This gives us a state space reduction property, where we can fully determine Ω_t from Ω_{t-1} and z_{t-1} by a recursive sequence, where the initial value is $\Omega_0(z^0) = 1 - F(z_0)$.

From here, we can substitute back into the FOCs:

$$(1 + \kappa) \left[\frac{\partial U_{t-1}}{\partial \pi_t(z^t)} f(z^{t-1}) + \beta \frac{\partial U_t}{\partial \pi_t(z^t)} f(z^t) \right] = \kappa \left[\Omega_{t-1}(z^{t-1}) \frac{\partial^2 U_{t-1}}{\partial z_{t-1} \partial \pi_t(z^t)} + \beta \Omega_t(z^t) \frac{\partial^2 U_t}{\partial z_t \partial \pi_t(z^t)} \right]$$

From here, it is helpful to divide through by $f(z^{t-1})$:

$$(1 + \kappa) \left[\frac{\partial U_{t-1}}{\partial \pi_t(z^t)} + \beta \frac{\partial U_t}{\partial \pi_t(z^t)} f(z_t | z_{t-1}) \right]$$

$$= \kappa \left[\frac{\Omega_{t-1}(z^{t-1})}{f(z^{t-1})} \frac{\partial^2 U_{t-1}}{\partial z_{t-1} \partial \pi_t(z^t)} + \beta \frac{\Omega_t(z^t)}{f(z^t)} \frac{\partial^2 U_t}{\partial z_t \partial \pi_t(z^t)} f(z_t | z_{t-1}) \right]$$

And from here, we define $\Gamma_t(z^t) = \frac{\Omega_t(z^t)}{f(z^t)}$. Note that we have:

$$\Gamma_t(z^t) = \frac{\Omega_t(z^t)}{f(z^t)} = \frac{\Omega_{t-1}(z^{t-1})}{f(z^t)} \frac{\int_{z_t}^{\bar{\theta}} \frac{\partial f_k(\theta_t | z_{t-1})}{\partial z_k} d\theta_t}{f(z_t | z_{t-1})} = \Gamma_{t-1}(z^{t-1}) \frac{\int_{z_t}^{\bar{\theta}} \frac{\partial f_k(\theta_t | z_{t-1})}{\partial z_k} d\theta_t}{f(z_t | z_{t-1})}$$

Giving us our key result for $t \geq 1$.

Note that the relevant initial condition is $\Gamma_0 = \frac{1-F(z_0)}{f(z_0)}$. This is the standard term in evaluating the virtual value in static mechanism design problems, and it is not surprising that it appears here. What is notable is that this term appears in the *date 1* optimality condition, in addition (as we will see) to the date-0 one. This is because of the time consistency problem.

Lastly, we can evaluate the FOC for π_0 . In π_0 , there is no time consistency element, and we are left with the simple tradeoff between current π and transfers. Repeating the steps from above, we obtain the simple condition

$$\frac{\partial U_0}{\partial \pi_0} = \frac{\kappa}{1 + \kappa} \Gamma_0(z^0) \frac{\partial^2 U_0}{\partial z_0 \partial \pi_0}$$

which is a standard virtual value condition. This gives the full result.

B.1.5 Proof of Corollary 9

The proof follows immediately from Proposition 8, noting that the triple can be used to evaluate the allocation rule from date t and onward, and that the transfer rule can be defined from the value function.

B.1.6 Proof of Corollary 10

The proof follows immediately from the definition of Γ_t , which is equal to zero if $\theta_t \in \{\underline{\theta}, \bar{\theta}\}$. When $\Gamma_t = 0$, the allocation rule is constrained efficient for all Γ_{t+k} , $k \geq 1$, so the optimal mechanism reverts to constrained efficiency, which is implemented by the dynamic inflation

target.

B.1.7 Proof of Proposition 11

The optimization problem

$$\max_{\{\pi_t\}} E_{-1} \sum_{t=0}^{\infty} \beta^t (1 + \kappa) U_t$$

yields the constrained efficient allocation rule. The result of Proposition 4 gives the mechanism that implements this allocation rule, giving the result.