Title
Three Essays on Computationally Intensive Economic Problems

Author
Lee, Seung

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California

Santa Cruz

Three Essays on Computationally Intensive Economic Problems

A dissertation submitted in partial satisfaction

of the requirements for the degree

DOCTOR OF PHILOSOPHY

in

Economics

by

Seung Lee

December 2016

The Dissertation of Seung Lee is approved:

Professor Kenneth Kletzer, Chair

Professor Eric Aldrich

Professor Grace Gu

Tyrus Miller

Vice Provost and Dean of Graduate Studies

Contents

List of Figures

1 Acknowledgments

I would like to thank the Economics Department at University of California Santa Cruz. I am grateful for their support. I would like to especially thank my committee: Professor Kenneth Kletzer, Professor Eric Aldrich and Professor Grace Gu for giving me invaluable knowledge and time.

I would also like to thank Professor Lilia Maliar and Professor Serguei Maliar for their support and knowledge.

I am forever indebted for everyone's help during my time at University of California Santa Cruz.

Three Essays on Computationally Intensive Economic Problems

by Seung Lee

2 Abstract

The chapters consist of three essays computationally intensive economic problems. In my first chapter, I model an economy, which decides to default by collective action. The model incorporates agents of various age and wealth to determine whether the country should default on their debt. In my second chapter, we explore the impact of market width in informational impact between exchanges. Using high frequency data, we show market width plays an important factor in determining order flow from one exchange to another. Final chapter explores state of Social Security Trust Fund. The model used to project Social Security uses up to date education and survival probabilities to emphasize the economy will be quite different in the future.

Chapter 1: Vote to Default

Abstract

Inspired by the political events that followed after the sovereign debt crisis in Greece post 2009, I develop an overlapping generations model with aggregate and idiosyncratic shocks to analyze agent's decisions if each had a vote in whether the country should default or not. The hypothetical economy where agents vote to default almost became a reality in 2015 when Prime Minister of Greece asked the voting population whether the country should remain in the bailout program through a referendum. Model results exhibit similar patterns as the Greek referendum with the young and less wealthy more inclined to vote for default.

Keywords: OLG, sovereign default, Krusell-Smith.

3 Introduction

In this chapter, I will present a model in which a country decides to default by collective action. Since the work of Eaton and Gersovitz (1981) sovereign defaults are modeled using a representative agent who defaults if the current and future streams of utility of default is larger than the case of not defaulting. The model I present replaces the representative agent with a population that is heterogeneous in dimensions of age, wealth and productivity. The heterogeneity in multiple dimensions allow for a richer analysis of the voting outcome than in unitary dimension, where voting outcome is determined by identifiable median voter.

The theoretical model is inspired by Greek elections of 2012. In the election, clear divide occurred between pro-bailout and anti-bailout parties. The pro-bailout parties supported the guidelines set forth by the creditors for new loans needed to service previous debt. The pro-bailout is a misnomer because these parties favored debt repayments in a newly structured manner. They wanted to repay debt in order to maintain the borrower-lender relationship. The anti-bailout parties wanted to reject the terms offered by the creditors as they did not believe the consequences to be more severe than the terms of bailout. Greek debt crisis was more complicated than debt repayment. Other issues included continued membership in the European Union, use of common currency in the Euro-zone, capital flight and health of Greek banking institutions. However the central election issue seemed to be debt and the austerity measures, set forth by the creditors. The citizens of Greece were indirectly voting on whether the country should default or not in the elections of 2012. In 2015, Greece debt crisis was still ongoing and impasse between Greece and creditors resulted in a referendum, in which Greek citizens were asked if they should accept the terms for continued bailout. This referendum was highly unusual in that the the population would be voting in a matter usually reserved for elected politicians and finance ministers.

The Greek debt crisis took center stage in Greek politics. When elected policy makers have the authority to decide whether a country repays their debt, the decision process is as much political as economic. The literature recognizes the importance of political influence in sovereign defaults. Borensztein and Panizza (2008) document the possible political costs associated with defaults. They find the executive branch turn over is twice as high after a default. They find similar results for finance ministers after a default. Livshits et al. (2014) find similar results but they are unable to statistically conclude defaults are associated with higher executive turnover.[1] Both studies imply the end of tenure for the current finance ministers when defaults occur during their office. The following section will present some models that try to incorporate a political element to sovereign default decision.

Aguiar and Amador (2011) incorporates political element to sovereign default by adding political turnover to the default decision. The incumbent party prefers higher consumption during their incumbency therefore discount the future by the discount rate and the probability of winning re-election, which is exogenous. Cuadra and Sapriza (2008) develop a two party model within the framework of Eaton and Gersovitz (1981). The two parties represent two different types of representative agents in the model. As in Aguiar and Amador (2011) political turnover occurs with exogenous probability. The party in office tries to maximize the utility of the population, weighting one group more than the other. In both models, the political uncertainty and preference to increase consumption during incumbency sustain debt and produce defaults.

Guembel and Sussman (2009) formulate a model in which the sovereign cannot discriminate between domestic and foreign lenders. In their setup, the median voter prefers debt enforcement, therefore positive foreign debt can be sustained in the absence of any punishment. In contrast to the previous models, the model provides an endogenous political process. The main element, however, is the inability to discriminate between foreign and domestic debt holders. This idea is formalized in Broner et al. (2010). In this model, the governments debts are traded on secondary exchanges in which domestic and foreign citizens can purchase/sell original debt. In their model, there is no political process, but the model demonstrates that the inability to discriminate between foreign and domestic lenders can sustain positive

[1]Note in most models, where politicians maximize length of office, tying defaults to office turnover generate no or few defaults since the cost of default is too high.

foreign debt.

The heterogeneity in the previous models occurred with respect to parties, individual types or idiosyncratic shocks they faced. Tabellini (1991) models domestic debt distribution within a two-period overlapping generation model. Similar to social security, government supplements the income of old with taxes on wage receipts and debt. The young and the old vote on a tax rate. In his model, the young with inheritances and wealthy elders form a coalition to repay debts. The model highlights the different objectives of the young and old, as old wish to maximize the wage tax in order to collect higher pension and the young wish to minimize their tax burden.

My model will try to explore the inter-generational conflict as in Tabellini (1991) in the context of sovereign debt. The Greek referendum of 2015 showed demographics played a key role in the vote. Significantly higher proportion of younger and/or less wealthy voted for NO. The model will be used to explore the effect of age and wealth in an individual's decision to vote for default. Numerical simulations generate similar the inter-generational and wealth divide as in the referendum results.

There are significant technical challenges in extending an overlapping generation model to sovereign defaults. A key element of sovereign default model is the aggregate economy stochastic process. Aggregate stochastic process inclusion in an overlapping generation model is non trivial. The state space is enlarged to include aggregate and individual capital, age, and stochastic processes. The constraints such as the participation and bond price constraints are difficult to parametrize even with high order polynomials. The solution method should be computationally efficient in order to traverse through a such large state space. The paper will be accompanied by efficient codes, which hopefully other researchers can modify to further increase the complexity of economic models as more advanced computing resources become available within a desktop environment.

Tabellini (1991) and Guembel and Sussman (2009) explicitly use the median voter theorem to analyze the models. The median voter theorem developed by Black (1948) is often used in analysis of public choice models. If the median voter exists, then the median voter will decide the policy. However the existence of the median voter is not guaranteed. It is not certain the median voter exists, especially if election turnout is less than 100%. It is also not useful when policy dimension is more than one. In the two dimensional case, median voter can only exist if the population distribution is perfectly symmetric, McKelvey (1979). Instead of trying to identify the median voter, the vote to default model will sum the votes of the created population.

The layout of the paper is as follow. I will detail the Greek debt crisis and the events that led to the referendum of 2015. Then I will survey the sovereign default and relevant heterogeneous economy literature. I will present few numerical exercises of representative agent sovereign default models as a comparison to the vote to default model. Then the model will be presented in detail along with numerical solutions of the model.

4 Greek Crisis

In October of 2009, newly elected Prime Minister Papandreou revealed Greece had been under reporting deficits for years in the backdrop of the Great Recession. The news came about as Greek economy was worsening as they would report GDP loss of 4.2% in 2009 after a loss of 0.3% in 2008. As investors fled Greek bonds, the yields rose sharply (900 basis points over German Bund, Zettelmeyer et al. (2013)). Greece was one of the more indebted nations with debt to GDP over 100%. As the yields rose, it became impractical to finance their budget deficits through the international bond markets. Also debt payments from previous loans were mounting. Greece did not have these funds and sovereign debt crisis was imminent. In April of 2010, Greek government requested financial assistance from the European Union and IMF to meet their debt obligations.

In May of 2010, European Commission, European Central Bank and IMF collectively known as the *troika*, agreed to a rescue package conditional on economic reforms. These conditional economic reforms also known as austerity measures were highly unpopular. Each austerity measure that passed through legislation reduced benefits such as pension and overtime pay and increased taxes from VAT to corporate. These measures were intended to reduce the budget deficits by focusing on expenditure cuts but with the the continued recession, Greece continuously faced budget shortfalls. The economy worsened and Greece needed debt relief once again. Early 2012, Greece ''restructured" their debt with private investors taking haircut losses of approximately 60% Zettelmeyer et al. (2013). The private bondholders of Greek debt suffered heavy losses as they lowered the Greek debt obligations by more than €100 billion. The bond restructuring and further aid from the *troika* required Greece to continue to implement austerity measures.

The debt crisis took center stage in Greek politics. On May of 2012 Greece held an election for all the seats in their congress. The austerity measures were the focus of the election. Some parties such as the current majority party New Democracy were in favor of austerity measures. Syriza, holding only 13 of 300 seats pre-election campaigned against austerity measures. Post election, they gained 39 seats and became the major opposition party. This election highlighted the close link between sovereign default and politics, which would become further evident in the events to come.

By 2014, Greek economy improved, reporting positive GDP growth of 0.65%. Greece even returned to the bond market raising over €6 billion from sale of bonds. Still the government needed tranches from the previous bailouts to finance their budget deficits. The *troika* and Greece started negotiations on Greece's continued implementation of the bailout program as elections were looming.

In January 2015, Greece held an election for all the seats in their congress. The central focus of the elections was the bailout. Some political parties including the current majority New Democracy argued for completing the bailout program. Anti-bailout parties such as Syriza argued to exit the the bailout program and force the *troika* for better terms of the bailout. Syriza, which came into prominence in the election of 2012 won 149 out of 300 seats, nearly capturing absolute majority. Their party leader Alexis Tsipras, who was a prominent voice in the anti-bailout movement became the new Prime Minister.

Greece faced a heavy schedule of debt repayments in early 2015. Without further austerity measures, tranches from earlier bailouts would not be released. Without these tranches, Greece would be the first country to default on IMF debt obligations. Unable to come to agreement with creditors, Greek Prime Minister Alexis Tsipras announced a referendum in June 2015. The referendum asked whether Greece should approve the proposal by the Juncker Commission, IMF, and ECB in regards to future of the Greece's bailout program. The proposal outlined the necessary initiatives, such as pension and public wage cuts, Greece needed to implement for further aid. There were many questions in regards to the legality of the referendum and accuracy of the referendum question as negotiations between the *troika* and Greece were ongoing. The consequences of ''No" on referendum was unclear. Tsipras who recommended a ''No" vote believed Greece can force *troika* to relax their austerity measure requirements by leaving the current bailout program. Others including the European Commission interpreted the referendum as whether Greece wanted to remain in the Eurozone. Although the question was not a direct question on whether Greece should default, the public was asked to decide on a matter usually reserved for high level ministers.

The outcome of the referendum was ''No" as it captured 61.31% of the votes. Soon after the referendum results, Greece proposed to the *troika* a bailout program that required less austerity measures. A week after the referendum Greece agreed to a bailout package that included larger tax increases and pension cuts than the one proposed by the Juncker Commission but future bailout loans included longer payment periods and lower interest rate. The ''No" referendum did not result in outright sovereign default nor exit of Greece from Europe as many feared. Given the post referendum results, the referendum was a negotiating tool and whether it was successful or not is unclear. However the referendum was a natural experiment on how different demographics saw the bailout issue. Although the implications of ''No" were unclear, the question posed had strong overtones of whether Greece should exit the bailout or not.

The referendum results are displayed in Figure 1. The referendum showed a divide between the young and old voters. The younger voting groups heavily favored the ''No" vote and negative relationship between age and ''No" is apparent. The oldest age group 65 and over favored the ''Yes". This seems odd since ''No" was a rejection for austerity measures such as pension reductions and increase in retirement age. The older age groups should favor policies that maximizes the pension, which at the time the ''No" referendum was more likely of the two to do so. Second there is negative relationship between wealth and ''No" vote. People facing financial difficulty voted 63% in favor of ''No" while 52.3% of the people who were living comfortably did. If the referendum was to hurt capital owners in terms of capital flows in and out of Greece or even exit from Europe as some feared, then it would seem rational that capital holders vote for the status quo, which would have been a''Yes" on the referendum.

The rise of the Syriza, Prime Minister Tsipras, and the referedum reflect the political polarization between the young and the old. Figure 2 displays the results of 2012 election when Syriza rose to prominence. Syriza's appeal to the younger voters are clearly apparent from the table. The results of the 2012 elections and 2015 referendum support the inter-generational conflict in Greek politics. If demographics matter in politics then it should also be a factor in sovereign defaults when sovereign defaults often become politicized. The paper will explore the economic effects of age and wealth in an agent's decision to favor sovereign default or not.

Greek Referendum 2015 : “NO” voter demographics

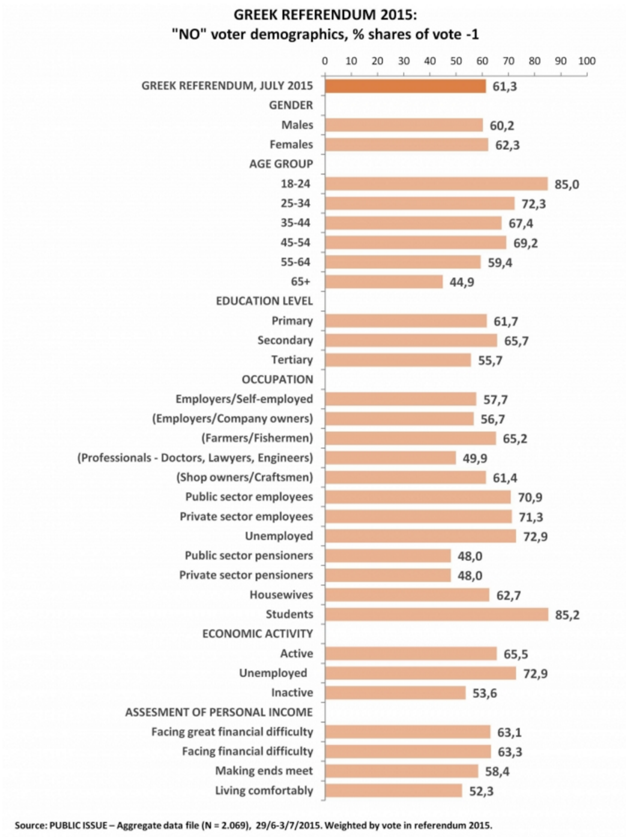

Figure 1: Referendum Results

Source: Public Issue - Total data set of the Pre-electoral Political Barometer (N=5.862 individuals) - May - June 2012

Voter Demographics:

	ND	SYRIZA	PASOK	IND. GREEKS	GOLDEN DAWN	DEM. LEFT	KKE	CREATION AGAIN	LAOS	OTHER
GREEK ELECTIONS 6/2012	29,7	26,9	12,3	7,5	6,9	6,3	4,5	1,6	1,6	2,7
GENDER										
Males	30	25	14	6	10	5	5	2	2	3
Females	30	29	11	9	4	8	4	1	2	3
AGE GROUP										
18-24	11	37	5	7	13	10	5	4	2	6
25-34	16	33	6	10	16	5	4	3	1	6
35-44	21	32	7	10	11	7	4	2	3	4
45-54	24	34	9	8	7	7	5	2	1	2
55-64	33	27	14	7	4	6	5	1	1	2
65++	48	13	21	5	2	5	4	1	2	1
EDUCATION LEVEL										
Primary	42	19	21	7	3	4	5	0	1	1
Secondary	27	29	11	8	9	5	5	1	3	3
Tertiary	28	28	11	7	6	9	4	3	1	4
EMPLOYMENT STATUS										
Employers / Self-employed	28	27	10	8	11	5	4	3	2	3
Public Sector salaried employees	21	33	10	8	6	10	5	1	1	5
Private Sector salaried employees	19	34	7	8	11	7	5	2	1	4
Unemployed	17	37	5	11	12	6	4	1	2	4
Pensioners	42	18	20	5	3	5	4	1	1	1
Housewives	37	23	11	9	3	6	5	1	3	3
Students	12	39	8	8	7	11	4	4	4	6

Figure 2: Referendum Results

5 Related Literature

5.1 Sovereign Default

The sovereign default decision at the core is a participation constraint. In the work of Eaton and Gersovitz (1981) defaults occur if the value of autarky is greater than the value of being in good standing with the creditors. The model is the workhorse of many sovereign default models. Their model in recursive form is:

$$V^D(y_t^a) = \mathbb{E}\sum_{t=0}^{\infty}\beta^t U(y_t^a) \tag{1}$$

$$V^R(y_t, d_t) = \max_{d_{t+1}} U(y_t + d_{t+1} - (1+r)d_t) + \beta\mathbb{E}\; max[V^R(y_{t+1}, d_{t+1}), V^D(y_{t+1}^a)] \tag{2}$$

Equation 1 is the value of defaulting. Under autarky the value of default is the infinite sum of expected discounted utility of autarky endowments denoted as y_t^a. Endowments stream are assumed to follow a autoregressive stochastic process. Utility is derived from consumption, which under autarky is just the endowment each period. Equation 2 is the value of staying in the borrower/lender relationship. In an endowment economy, the value of being in good standing has two advantages. One because there is no method for savings in the domestic economy, the loans facilitate consumption smoothing. New loans denoted as d_{t+1} minus previous debt denoted d_t with added interest can augment consumption under no default. Second, the production or endowment under good standing is higher than under default. The punishment for default is $y_t^a < y_t$. The drop off in production may occur in the form of trade retaliation. Bulow and Rogoff (1989) include trade sanctions in the Eaton and Gersovitz (1981) framework to incorporate a more severe punishment than reputation costs alone.

Although lower output is associated with defaults, it is unclear whether defaults lower output for defaulters. In fact, it is even uncertain whether defaults occur during lower output periods according to Tomz and Wright (2007). They reveal in their study only 60% of defaulting countries in their sample defaulted when output was below trend[2]. Levy-yeyati and Panizza (2006) looked at the quarterly growth of defaulting countries and they find periods immediately after default exhibit positive economic growth. The loss in output precedes the default and the trough of the contraction occurs at the default episode. Empirical evidence cast doubt whether defaulting countries suffer from output due to act of default. However it is common in theoretical literature to include a punishment clause for defaulting. Without the clause, the models have a difficult time generating defaults under simulation.

Beginning of each period, an endowment is realized from a stochastic process. After the realization of the endowment, the borrower decides to participate in the borrower/lender relationship if participation constraint $V^R(\cdot) \geq V^D(\cdot)$ is satisfied, which states that the value of paying off debt has to be better than autarky. The value of staying in the borrower/lender relationship and value of default along with the constraint fully characterize the model.

The first models to extend this framework to match the business cycle were Arellano (2008) and Aguiar and Gopinath (2006). I will focus on the Arellano (2008) as the two models are very similar. Arellano (2008) is able to match many empirical moments by adding a pricing kernel and endogenous autarky value.

$$V^o(y_t, d_t) = max[V^R(y_t, d_t), V^D(y_t^a)] \tag{3}$$

$$V^D(y_t) = U(y_t^a) + \beta\mathbb{E}[\theta V^R(y_{t+1}, 0) + (1-\theta)V^D(y_{t+1}^a)] \tag{4}$$

$$V^R(y_t, d_t) = \max_{d_{t+1}} U(y_t + q(d_{t+1}, y_t)d_{t+1} - d_t) + \beta\mathbb{E}V^o(y_{t+1}, d_{t+1}), \tag{5}$$

where θ is exogenous probability of re-entering the lending markets and $q(\cdot)$ is the price of bonds.

Equation 4 is the analog of equation 1, except here the lender has some exogenous probability of re-entering the debt markets with zero debt. The exogenous re-entry is used because defaulting countries return to debt markets soon after they default. Some models such as Yue (2010) replace the exogenous re-entry with debt renegotiation between creditors and debtors in a Nash bargaining setting. Kletzer

[2]Their data was HP filtered

and Wright (2000) show how the borrower lender relationship is sustainable even without penalties and exogenous re-entry if both parties have limited commitment. In their model the borrower lender relationship is characterized by consumption smoothing. The lender pays the borrower in low endowment periods and the borrower repays during high endowment periods. If the borrower defaults and tries to re-establish such a state contingent relationship with a new lender, the original lender has the option to "cheat the cheater". The model illustrates how punishment does not require cooperation among the lenders.

Equation 5 is the analog of equation 2. The innovation here is that by using a pricing kernel of bonds $q(d_{t+1}, y_t)$, she is able to construct a Laffer curve for which the price of bonds drop as country incurs higher debt.[3] The bond prices satisfy the zero profit conditions of the lenders that is they do not earn more than the risk free rate in expectations. Price of the bonds satisfies

$$q(d_{t+1}, y_t) = \int_{y \in def} \frac{1 - f(y)dy}{1 + rf}, \tag{6}$$

where $f(y)$ is the probability density function of the endowment process.

The bond prices are dependent on the amount the country will borrow and the expected shock next period. The In period 1, the country decides amount to borrow. In period 2, the country decides whether to repay the borrowed amount in period 1 taking into account the endowment realization of period 2. Therefore the bond pricing is a function of the amount the borrower this period and the expected endowment realization for next period. Every endowment realization that results in default decreases the price of the bonds by the probability of that realization, normalized by the risk free rate. If there are no events that cause defaults then the maximal price of the bonds is the inverse of the risk free rate. The borrower in Eaton and Gersovitz (1981) received 1 unit for every bond unit they borrowed and repaid the unit of bond with added interest. Arellano (2008) uses discounted one period bonds with a face value of one unit. The Arellano (2008) model will be the basis for the government problem in the vote to default model.

[3] Laffer curve here refers to the notional amount the government can borrow $q(\cdot) \times d_{t+1}$.

5.2 Heterogeneous Agents

In the work , Auerbach and Kotlikoff (1987) fiscal policies are analyzed using a multi period overlapping generation model. In an overlapping generation model, each generation has their own budget constraint. Each of these are dynamically linked by capital holdings that carries over each period i.e. an agent carries over to next period, what she saves this period. The model equations are

$$\max_{k_{t+1}^1, k_{t+1}^2, \dots, k_{t+1}^{G-1}} \sum_{s=1}^{s=G} \beta^{s-1} U(c_t^s) \tag{7}$$

$$k_{t+1}^{s+1} = (1 + r_t - \delta) k_t^s + (1 - \tau_t) w_t n_t^s - c_t^s, \quad s = 1, 2, \dots R - 1 \tag{8}$$

$$k_{t+1}^{s+1} = (1 + r_t - \delta) k_t^s - c_t^s + pen_t, \quad s = R, R+1, \dots G - 1 \tag{9}$$

$$k_{t+1}^G = 0 \tag{10}$$

$$K_t = \sum_{s=1}^{G} k_t^s \tag{11}$$

$$N_t = \sum_{s=1}^{R-1} n_t^s \tag{12}$$

$$C_t = \sum_{s=1}^{G} c_t^s \tag{13}$$

$$K_{t+1} = (1 + r_t - \delta) K_t + w_t N_t - C_t \tag{14}$$

$$r_t = \alpha \left(\frac{K_t}{N_t} \right)^{\alpha - 1} \tag{15}$$

$$w_t = (1 - \alpha) \left(\frac{K_t}{N_t} \right)^{\alpha} \tag{16}$$

An agent maximizes their lifetime utility, which is the utility from age 1 to age G. During the work years, which is age 1 to $R - 1$ they have wage income. During their retirement years, they no longer work but collect a pension. Once the agents reach the age G, they do not save anything since they will not survive beyond this age. Equations 11-14 are the market clearing conditions. Total capital in the economy is the sum of each generations' capital holdings and total labor in the economy is the sum of the labor hours provided by the generations that work. Market prices r_t and w_t are factor prices for Cobb Douglas production of aggregate holdings of capital and labor. Auerbach and Kotlikoff (1987) model naturally creates a wealth distribution with respect to age as agents save for their retirement. In these types of models, a hump shaped age-wealth profile is generated as agents start and end with 0 wealth with the peak occurring at at age of retirement.

The model of Auerbach and Kotlikoff (1987) is deterministic. The vote to default model will have two stochastic processes: idiosyncratic and aggregate. Models with idiosyncratic shocks create a wealth distribution that must be accounted when determining the aggregate capital in the economy. In an economy with infinitely lived agent with either different endowment streams as in Huggett (1993) or with different productivity as in Aiyagari (1995), the distribution of wealth converges to a stationary distribution and the aggregation occurs with respect to the stationary distribution of wealth.

$$K_t = \sum_{i=1}^{2} \int_0^{\infty} af(i,a)da \tag{17}$$

Equation 17 is the analog of the equation 11 in Auerbach and Kotlikoff (1987). The distribution is with respect to individual type denoted i, which can be employment status or productivity type and a, which is her savings or wealth. For example assume half of the agents are of high productivity and the other half are of low productivity. Both agents face a idiosyncratic productivity shock that can be either high or low with equal probability. If in period 0, both types of agents start with no capital then in period 1, quarter of the population are of high productivity and high shock, quarter are high productivity and low shock, quarter are low productivity and high shock, and quarter are low productivity and low shock. If each of the quarter save a different amount, then we would have 4 distinct capital holdings. After sufficient iteration, stationary wealth distribution forms.

In order to apply a heterogeneous agents in a sovereign default setting, aggregate shocks are needed. If state contingent claims exist, there would be no need for foreign bonds to smooth consumption without the presence of aggregate shocks[4]. Krusell and Smith (1998) numerically solve a model with unemployment shocks and

[4]With lower discount factor than the world, foreign bonds will be demanded to tilt consumption forward but would not generate defaults

aggregate shocks in an infinite period model. The difficulty of including a stochastic aggregate process involves calculating next period's factor prices, which are functions of next period's capital and shock. The law of motion for aggregate capital stock becomes a state and is endogenous. In their model, they are able to use a linear law of motion for aggregate capital, in which agents' capital savings are consistent with the evolution of aggregate capital. Heer and Maußner (2011) apply the Krusell and Smith (1998) algorithm to an overlapping generations with idiosyncratic shocks and aggregate shocks. Their model is the basis of the household problem in the vote to default.

The models of Arellano (2008) and Heer and Maußner (2011) are computationally expensive. Models with participation constraint are not easily characterized by the first order conditions. Also note equation 3, which is the maximum of the default and no default will be a kinked function. Accurate numerical approximations of kinked functions are difficult and the difference between value of default and no default are quite small analogous to welfare loss being small in real business cycle models. OLG models are also computationally challenging. Solutions to OLG model entail solving difference equations in the order of the number of generations in the model. Finally the stochastic elements included in the model augment the already enlarged state space. The vote to default model will try overcome these computational difficulties.

6 Numerical Exercises

6.1 Endowment Economy

The model of Arellano (2008) is the basis for the government problem in the vote to default model. I hope to bring clarity to my model by exploring her model in detail. In her model, the small open economy faces a bond pricing schedule that is decreasing in the amount of outstanding debt $\frac{dq}{dd} < 0$ and increasing in the endowment $\frac{dq}{dy} > 0$. If the lenders are restricted to zero profits, then the lenders will earn the risk free rate in expectations. The price of bonds for any given state will be the expected probability of no default normalized to the risk free rate. Recall the bond pricing kernel of equation 6:

$$q(d_{t+1}, y_t) = \int_{y \in def} \frac{1 - f(y)dy}{1 + rf_t},$$

where $f(y)$ is the probability distribution of the endowment process and the integral is taken with respect to the endowment states that are expected to result in defaults.

The country borrows during low endowment periods and and repays the amount next period. For every unit they borrow, they receive the price of the bond $0 \leq p \leq \frac{1}{1+rf}$ and repay the unit borrowed next period at $p = 1$. If they decide not to default they can choose to borrow again. Capital inflow occurs if they choose to borrow an amount greater than the amount repaid $d_t < q(d_{t+1}, y_t)d_{t+1}$.

Figure 5 shows the bond choices for given debt and high or low endowment. During low endowment periods, they are usually net borrowers. During high endowment periods, they borrow less than they did previous period resulting in capital outflows. As they carry forward larger debt balances, the bond price will fall (Figure 4). If the bond price is sufficiently low, even in low endowment periods, they will not be able to add to consumption today. If the drop in consumption is sufficiently large, the future value of staying in the relationship will not be enough to offset the consumption loss today, which will result in default. The decision to default trades increase in consumption with future consumption smoothing.

There are two interesting features in the model that I wish to highlight. First the law of motion for the stochastic process is very important in the default decision. The endowment process has to be volatile enough for the borrower to repay debts. In the case of constant endowment, there will be no need for future consumption smoothing. In this case, the borrower will borrow the maximum in first period and default. Second, the default region predominantly lies in the area of low endowment and high debt (Figure 3). Note also the defaults do not occur when the country has a positive balance, that is when they are lenders. As the endowment process becomes less auto-correlated defaults can occur during high endowment periods. In the case of i.i.d. endowments even when the endowment is highest and marginal utility has the least decrease with debt service, the borrower will default if sufficiently in debt. Figure 4 right, shows the bond pricing schedule for an i.i.d. endowment process. The bond prices reflect the incentive for default in high debt regions regardless of the current endowment state.

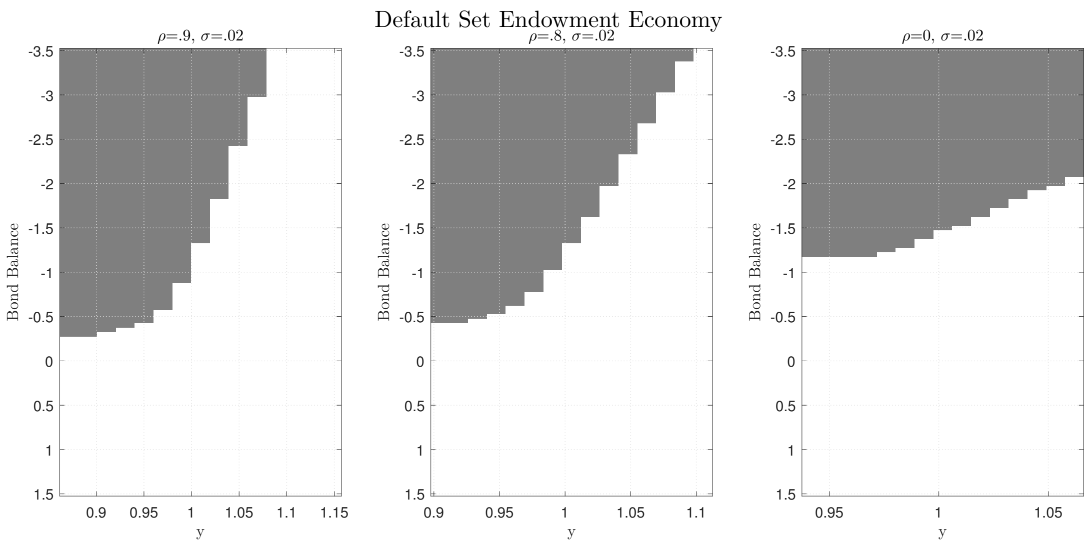

Figure 3: Default Set: Endowment Economy

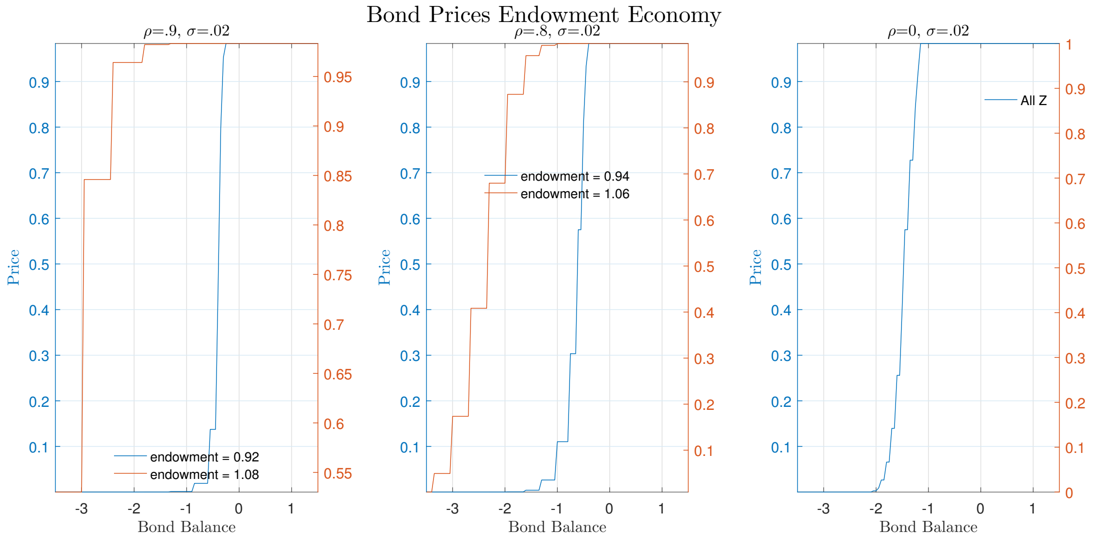

Figure 4: Price Schedule: Endowment Economy

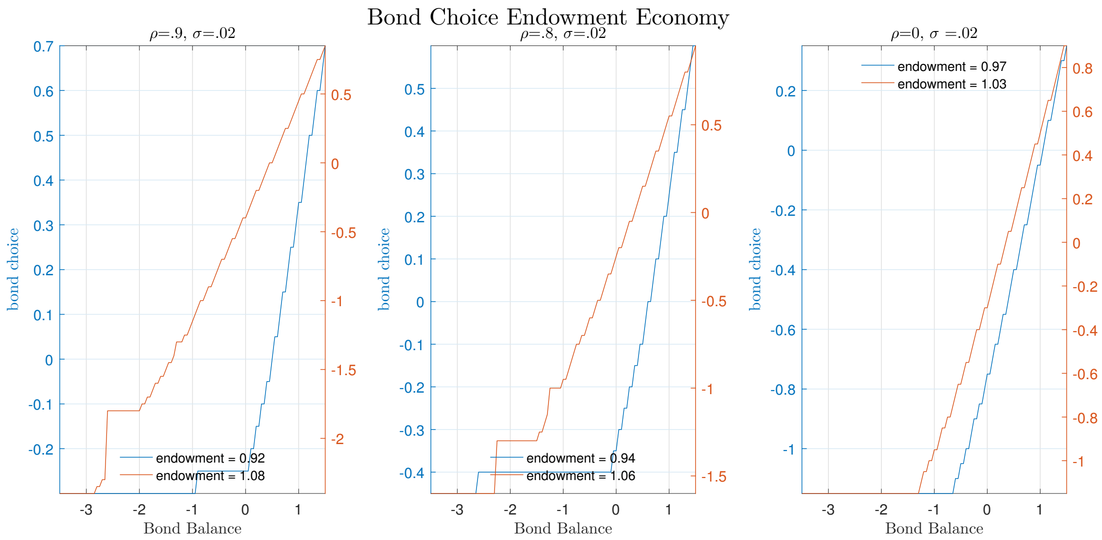

Figure 5: Bond Choice: Endowment Economy

6.2 Production Economy

In the production economy, capital stock adds a new state to the sovereign default problem. The value of default and no default are contingent on the capital stock today. The following equations summarize the model.

$$V_t^o(Z_t, k_t, d_t, k_t^a) = max[V_t^D(Z_t, k_t^a), V_t^R(Z_t, k_t, d_t)] \tag{18}$$

$$V_t^D(Z_t, k_t) = \max_{k_{t+1}^a} U(c_t^a) + \beta \mathrm{E}[(1-\theta)\ V_{t+1}^D(Z_{t+1}, k_{t+1}^a) + \theta V_{t+1}^R(Z_{t+1}, k_{t+1}, 0)] \tag{19}$$

$$V_t^R(Z_t, k_t, d_t) = \max_{b_{t+1}, k_{t+1}} U(c_t) + \beta \mathrm{E} V_{t+1}^o(Z_{t+1}, k_{t+1}, d_{t+1}, k_{t+1}^a) \tag{20}$$

$$c_t^a = (1-\tau_k)(r_t^a - \delta)k_t + (1-\tau_w)w_t^a n_t + k_t - k_{t+1}^a + g_t^a \tag{21}$$

$$c_t = (1-\tau_k)(r_t - \delta)k_t + (1-\tau_w)w_t n_t + k_t - k_{t+1} - d_t + q(k_{t+1}, d_{t+1}, z_{t+1})d_{t+1} + g_t \tag{22}$$

$$Y_t(Z_t, k_t, n_t) = Z_t k_t^\alpha n_t^{1-\alpha}; r_t = MPK; w_t = MPL \tag{23}$$

$$Y_t^a(Z_t, k_t, n_t) = \psi Z_t k_t^\alpha n_t^{1-\alpha}; r_t^a = MPK; w_t^a = MPL; \psi < 1 \tag{24}$$

$$Z_t = \phi Z_{t-1} + \sigma \nu_t; \nu_t \sim N(0,1) \tag{25}$$

$$\tau_k(r_t - \delta)k_t + \tau_w n_t w_t = g_t \tag{26}$$

$$\tau_k(r_t^a - \delta)k_t + \tau_w n_t w_t^a = g_t^a \tag{27}$$

$$q(d_{t+1}, Z_t, k_{t+1}) = \int_{Z \in def} \frac{1 - f(Z)dZ}{1 + rf_t}, \tag{28}$$

The value of default equation 19 now becomes a control problem as savings for next period needs to be chosen. The value of no default equation 20 has two controls, capital and debt. As before, default forces lower output for all time periods ($\psi < 1$ in equation 24). Autarky states and prices are denoted with superscript a. The capital choice for next period k_{t+1} will be different whether the country is in autarky or not as the law of motion for capital differ. Therefore when evaluating whether to default or not, optimal capital choices for default and no default will be needed to evaluate equations 19 and 20. Note both value of default and no default have the same state of capital stock today. What is different is the capital stock choice next period under default or no default. The bond pricing kernel is now dependent on three states: capital choice this period, bond choice for this period and the shock (equation 28). But as before, the bond price will reflect the probability of no default under chosen capital stock and bond, normalized to risk free rate.

The problem ignores labor choices and assumes the labor in inelastic. As in business cycle models, elastic labor supply will help propagate shocks. In the problem of sovereign default, the propagation of shocks increases the volatility of capital stock and adds an extra control to the model, which are both difficult to deal with in terms of computation. In order to make the model more tractable labor choice was simplified.

In the endowment economy, there existed no storage technology other than foreign bonds. In a production economy, the capital stock is the main source of consumption smoothing. Access to foreign bonds allows the small open economy to smooth their consumption in conjunction with capital savings and build their capital stock faster. Given lower time preference, the capital stock of the small open economy will be low enough to offer high rental returns. If the implied interest rate by the sovereign bonds are lower than the net rental return on capital, there will be opportunities for the country to use the available bonds to add to the capital stock. Equation 29 is the criteria in which the implied sovereign bond intrest rate is lower than the net capital rental return.

$$(1-\tau_k)(\alpha z_t k_t^{\alpha-1} n_t^{1-\alpha} - \delta) > \frac{1}{q(k_{t+1}, d_{t+1}, z_{t+1})}, \tag{29}$$

where τ_k is capital tax rate and δ is depreciation rate.

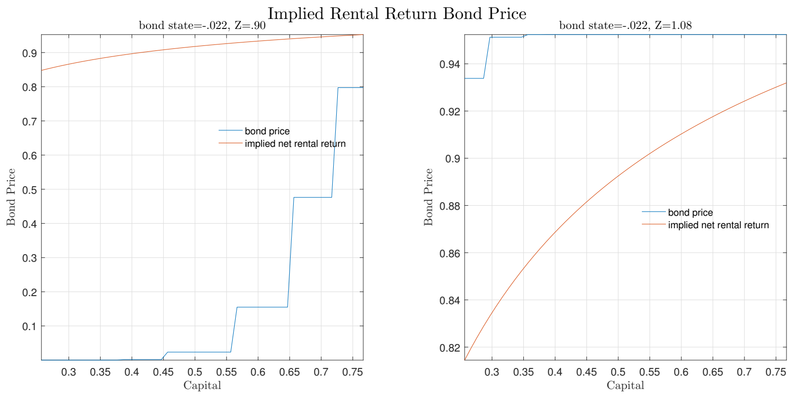

Figure 6: Investment Opportunities

Figure 6 shows the bonds pricing schedule (blue) for two cases, left low technology shock and right high technology shock. The bond prices implied by net capital rental rates are noted in red. In the case of high technology shock, the implied bond prices by net capital rental rates are lower than the sovereign bonds. In this case, the country can borrow and invest the loan proceeds in to their capital stock. Higher the gap between the implied bond prices of sovereign bonds and net capital rental rate, the probability of the investment yielding profits increases. Increases in the capital stock will increase the implied bond prices as the capital returns are lower and fewer opportunities will exist for using bonds to invest in capital. The production economy has an an investment channel which the endowment economy lacked. Before the country borrowed during lower endowment periods (Figure 5). In a production economy, the country will borrow in high endowment periods as well given there are investment opportunities. Production also reduces the need for sovereign bonds to smooth consumption. As in the endowment economy, the law of motion for the stochastic process in this case for technology shock is important.

The default set given low debt is depicted in Figure 7 top row. In the case of low capital stock, the default set shrinks as technology shock become less auto-correlated. Low capital stock and high technology shock implies high rental returns but with higher auto-correlation, the reliance on sovereign bonds decrease as higher future production becomes more likely with higher auto-correlation. The ability to consumption smooth using capital stock makes it more attractive to increase to consumption today and forgo the debt service. When capital stock is lowest, consumption is so low, even small additions to consumption will increase marginal utility significantly to make default a more attractive option than servicing debt.

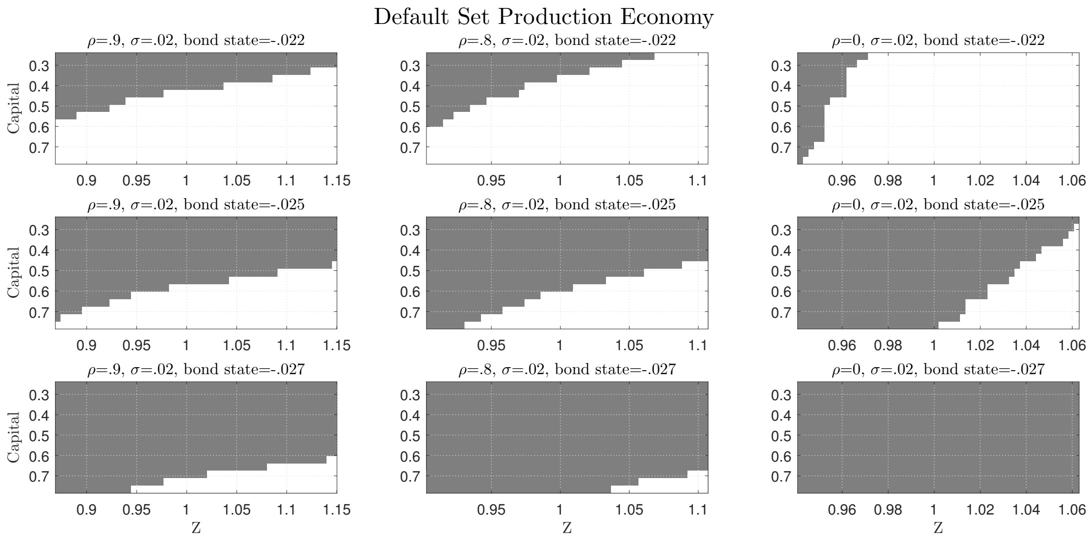

Figure 7: Default Set: Production Economy

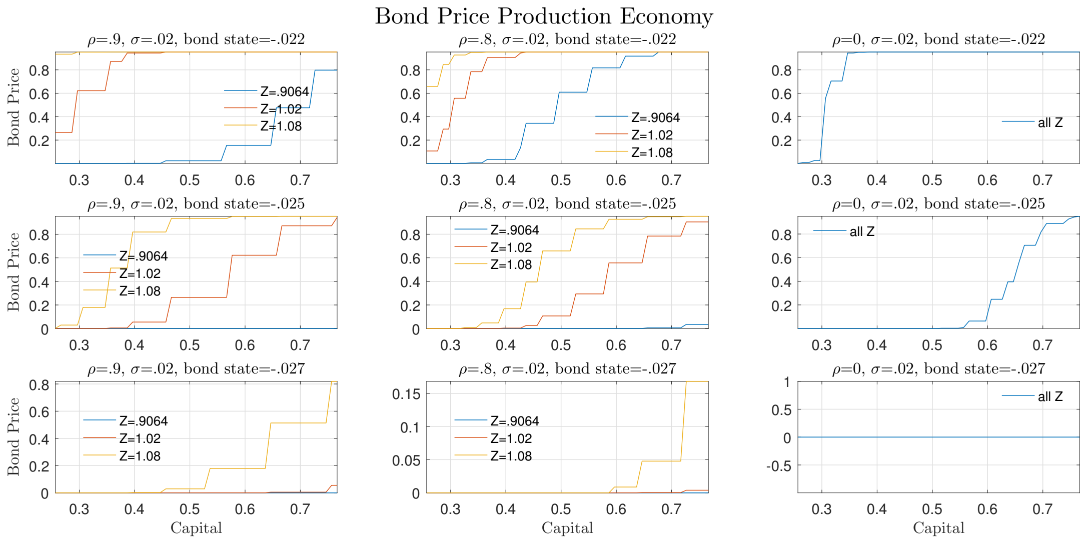

Figure 8: Bond Price: Production Economy

As capital stock increases, default set shrinks as the debt service has smaller effect on marginal utility today. But if the debt is too large, default option is better even with large capital stock. The pricing of the bonds reflect this as bond prices increases with capital stock (Figure 8). Bond choices when capital stock is large are smaller than when capital stock is small. Higher capital stock allows more consumption through production so less borrowing is needed. But as the technology shock increases, borrowing increases as bonds can be used for investment if rental returns are sufficiently high and high bond prices make cheap borrowing more attractive.

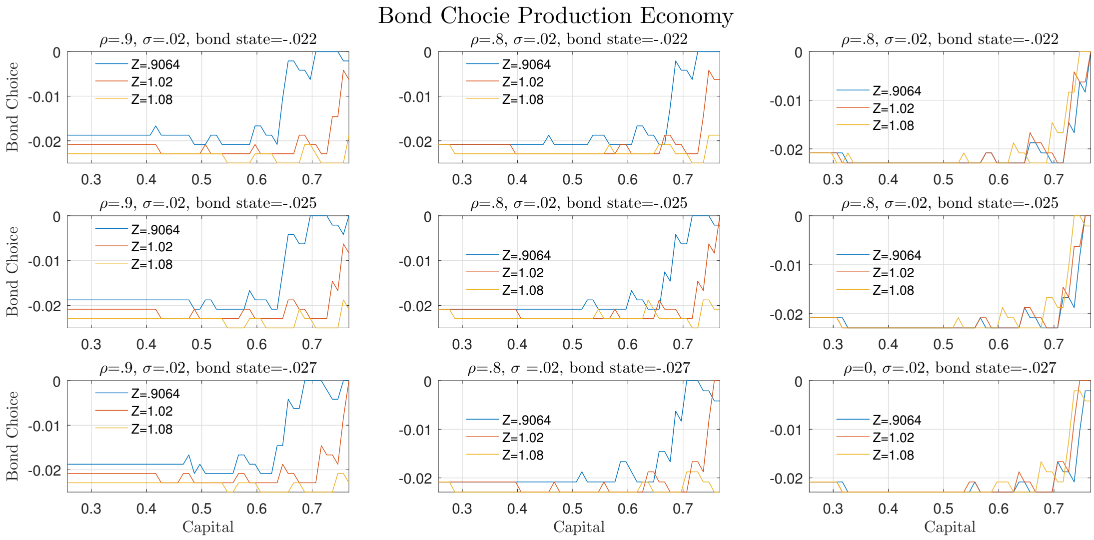

Figure 9: Bond Choice: Production Economy

7 Vote to Default Model

In this model, sovereign default is decided by collective action. In previous models, the government decided to default or not by comparing the value of default and no default for a representative agent. To create a non-degenerate distribution of votes, multiple agents will be needed. An overlapping generations framework will be used because the distribution with respect to age and wealth is of interest. In a static overlapping generations model, aging naturally creates a wealth distribution. Households accumulate capital for their retirement during the working years and deplete their savings during retirement. In order to create a more rich wealth distribution, Krusell and Smith (1998) variant of the overlapping generations model will be used.

There are high and low productive households. Each type faces a highly persistent idiosyncratic productivity shock that is either high or low. Each household lives up to a maximum age G[5]. In this economy, there is no population growth. Each year, unit mass of households are born. Each year the households face a survival probability of $\phi(s)$, which is age s dependent. For example the new born households

[5] In the model, G is set to 30.

survive to year 2 with probability $\phi(1)$[6]. The households will therefore have to discount their future value by the discount rate β and the probability of surviving to next age $\phi(s)$. There are three distinct phases of a household. First is the last age $s = G$. If the household has survived to age G, then the capital choice and the value next period regardless of default or not is 0 since they can not live past this age. Second is their retirement period excluding the last year of retirement. During the retirement period, their source of income are social security, which is funded by a tax on wages, return on their capital savings and distribution of net loan receipts from the government, which can be negative if there is capital outflow. The last phase is their working years, ages 1 to $R-1$. During the work years, their sources of income are wages, return on capital savings and distribution of net loan receipts from the government.

$$\begin{aligned} &V_t^{o,s}(Z_t, k_t^s, e_t, i, K_t, K_t^a, B_t) = \\ &\quad max\ [V_t^{D,s}(Z_t, k_t^s, e_t, i, K_t, K_t^a), V_t^{R,s}(Z_t, k_t^s, e_t, i, K_t, K_t^a, B_t)] \end{aligned} \tag{30}$$

Equation 30 is the analogous to equation 18 but there are additional state variables. Lower case variables denote individual states and upper case denote aggregate states. The value for default and no default have age s superscripts since these will be dependent on age. Individual capital stock also have age superscripts since these choices are age dependent. Idiosyncratic shock that is either high or low is denoted e_t. Household productive type that is either high or low is denoted i. Aggregate capital stocks are denoted K_t and K_t^a. Government's dispersion of net loan receipts is denoted $B_t = q(\cdot)D_{t+1} - D_t$.

$$\begin{aligned} V_t^{D,s}(Z_t, k_t^s, e_t, i, K_t, K_t^a) = \max_{k_{t+1}^{a,s+1}} U(c_t^{a,s}) + \beta \mathrm{E}[(1-\theta)\ V_{t+1}^{D,s+1}(Z_t, k_{t+1}^{s+1,a}, e_{t+1}, i, \\ K_{t+1}, K_{t+1}^a) + \theta V_{t+1}^{R,s+1}(Z_{t+1}, k_{t+1}^{s+1,a}, e_{t+1}, i, K_{t+1}, K_{t+1}^a, 0)] \end{aligned} \tag{31}$$

[6] Survival probability is inferred from the Social Security Administrations Actuarial Life Table for United States 2015.

$$V_t^{R,s}(Z_t, k_t^s, e_t, i, K_t, K_t^a, B_t) = \max_{k_{t+1}^{s+1}} U(c_t^s) + \beta \mathrm{E} V_{t+1}^{o,s+1}(Z_{t+1}, k_{t+1}^{s+1}, e_{t+1}, i, K_{t+1}, K_{t+1}^a, B_{t+1}) \tag{32}$$

$$c_t^{a,s} = (1-\tau_k)(r_t^a - \delta)k_t^s + (1-\tau_w)(1-\tau_s)w_t^a n_t \Gamma(i, \epsilon_t, s)\mathbb{1}_{s<R} + k_t^s - k_{t+1}^{s+1,a} + (1-\tau_k)ss_t^a \mathbb{1}_{s\geq R} \tag{33}$$

$$c_t^s = (1-\tau_k)(r_t - \delta)k_t^s + (1-\tau_w)(1-\tau_s)w_t n_t \Gamma(i, \epsilon_t, s)\mathbb{1}_{s<R} + k_t^s - k_{t+1}^{s+1} + B_t + (1-\tau_k)ss_t \mathbb{1}_{s\geq R} \tag{34}$$

Equations 31 and 32 are the values under default and no default respectively. The continuation value for both values have aggregate state variables that are of next period. Aggregate capital stock today is

$$K_t^a = \sum_{i=1}^{2}\sum_{s=1}^{G} k_{t,}^a(s,i) \tag{35}$$

$$K_t = \sum_{i=1}^{2}\sum_{s=1}^{G} k_t(s,i). \tag{36}$$

Aggregate resource constraints satisfy:

$$K_{t+1} = C_t + (1 + r_t - \delta)K_t + w_t N_t - B_t \mathbb{1}_{vote=nd} \tag{37}$$

The aggregate capital stock is the summation of the capital holdings by the age groups and productivity types. During the retirement years, households productivity can be ignored since they do not have wage income. The aggregate capital stock for next period is needed to decide the individual capital choices for next period. Therefore the households will need to know the law of motion for aggregate capital stock under default and no default. Krusell and Smith (1998) parametrize the law of motion for aggregate capital stock with a linear form using current state variables. For this problem, I use a lookup table of that is dependent of current K_t, Z_t and D_t. The lookup table or law of motion needs to be sufficiently accurate such that $K_{t+1} = \sum_{i=1}^{2}\sum_{s=1}^{G-1} k_{t+1}(s,i) \pm tolerance$ [7].

[7] Due to default decisions which are binary, it is difficult to characterize the aggregate law of motion linearly. If domain of law of motion for aggregate capital lies on a bounded real line, then

The continuation value for no default is dependent on future net bond receipts B_{t+1}. This is determined by the government bond policy rule that is dependent on K_t, Z_t, and D_t. This will be further detailed under government problem.

Equations 33 and 34 are the budget constraints under default and no default. Under autarky, the income flows are capital rental returns after capital tax, wage receipts after social security and labor tax if households are working ($s \leq R$) and social security disbursements if retired ($s > R$). The gamma function $\Gamma(i, e_t, s)$ adjusts wages for households age s, productivity type i and the idiosyncratic shock e_t. The gamma function exogenously creates a wage-age profile that is hump shaped to capture the hump shaped wage-age profile in the data. The gamma function is normalized to one such that the mean wage of the population is w_t, the marginal product of labor. Under no default the budget constraint includes a flow of net loan disbursements, which can be negative if the government is notionally borrowing less this period than the loan repayment for previous period's borrowing. The social security payments are such that the receipts from the workings are distributed to the the retired and balances every period. In equations 38 and 39, $mass(i, s)$ denote the population of productivity type i and age s.

$$ss_t \sum_{s=R}^{G} mass(i,s) = \sum_{i=1}^{2} \sum_{s=1}^{R-1} mass(i,s)\tau_s w_t n_t \tag{38}$$

$$ss_t^a \sum_{s=R}^{G} mass(i,s) = \sum_{i=1}^{2} \sum_{s=1}^{R-1} mass(i,s)\tau_s w_t^a n_t \tag{39}$$

The model economy is taxed heavily. The τ_s, τ_k, τ_w is set to .2, .3, and .3 respectively. Greece economy is also taxed heavily. Social security tax is 28.5% for employers and 16.5% for employees with a cap of €5546, the maximum tax with-holdings[8]. The average tax rate for personal income is 18%[9]. The high taxes in the economy are used in order to limit capital accumulation in the economy. These high tax rates reduce income to invest but also lower future rental returns

the default decision for the aggregate capital needs to be well characterized for this domain as well.

[8]https://www.oecd.org/ctp/tax-policy/Social-Security-Contributions-Explanatory-Annex-May-2015.pdf

[9]https://www.oecd.org/tax/revenue-statistics-greece.pdf

making investments less attractive. With lower taxes aggregate capital stock will increase/decrease significantly during high and low technology shocks respectively. This is problematic as volatile technology shocks are needed for default analysis. Taxation reduces the volatility of the aggregate capital stock, which allows for easier characterization for the law of motion for aggregate capital stock. A more suitable method for inhibiting capital growth is capital adjustment costs but I was unable to incorporate this feature into the model.

Given the states of the world, government bond redistribution plan and the future evolution of aggregate capital stock under default and no default evolve, the households will have the necessary information to decide whether their value is higher under default or no default.

The government problem is similar that of the production economy except the default decision is tabulated from votes. The government defaults if

$$\sum_{i=1}^{2}\sum_{s=1}^{G}\sum_{a=0}^{\bar{a}} mass(s,i,a) \in V^{D,s}(\cdot) > V^{R,s}(\cdot) > \frac{1}{2}, \tag{40}$$

where $\bar{a}$ is the upper limit of individual savings. If the mass of the population (total population is normalized to 1) who favors default, that is their value of default is greater than that of no default, is greater than half, then the government defaults. Under default, debt redistribution B_{t+1} is zero. If the vote favors no default then the government solves the following problem equation 41. Solution to the government's problem is the policy function that maps the amount to borrow for given states of the world: K_t, Z_t, and D_t. The policy function determines the amount the government wants to borrow to maximize stream of representative agent's utility. The net bond receipts B_t are of interest to the households which is $q(Z_{t+1}, K_{t+1}, D_{t+1})D_{t+1} - D_t$. The net bond receipts schedule is used by the households to evaluate the continuation value of no default $V_{t+1}^{r,s+1}$ which requires knowledge of next period's net bond receipts.

$$V_t^g(Z_t, K_t, D_t) = \underset{D_{t+1}}{max}\, U(y_t - K_{t+1} + (1-\delta)K_t - D_t + q(Z_{t+1}, K_{t+1}, D_{t+1})D_{t+1}) \\ + \beta \mathrm{E} V_{t+1}^g(Z_{t+1}, K_{t+1}, D_{t+1}) \tag{41}$$

The model mechanism is as follow. If the country is under good standing at time t, the agents in the model will have a choice to default or not. If the country has positive debt D_t that is they borrowed at time $t-1$, the country will hold an election to default or not. The agents in the economy incorporate the individual states: (wealth, productivity type, and idiosyncratic shock) and overall economy states: (aggregate capital stock this period, aggregate capital stock next period, aggregate shock today and aggregate shock tomorrow) to decide if they derive more value from defaulting or repaying debt.

The value of repaying debt is a function of maximum of value of repaying and value of defaulting next period because repaying debt allows you to default next period. The value of repaying debt has an embedded option to default. The value of defaulting or the punishment from default is lessened by the exogenous re-entry to the debt markets. Increasing the probability to re-enter the debt markets lowers the cost of default as autarky is lower output and re-entry guarantees zero debt along with embedded option to default again.

After the agents solve their problem and vote, the government tallies the votes. The government defaults if the vote to default is majority. The government budget is balanced. Any payment that would have been used to pay creditors will be used in terms of transfer payments. Another interpretation of the agent's decision to vote can be that they are voting to increase transfer payments. An interesting extension would be if instead of increasing transfer payments, the government lowered tax rates to balance the budget. This problem would be more in line with the Tabellini (1991).

Last point about the model before the results of the model are discussed. The agents in the model do not internalize their votes. The model prevents the agents to game the system by forming coalitions or to vote strategically. Although these are important and interesting components to study the model is incapable of including these features and find a solution in a practical time.

7.1 Equilibrium

The solution of heterogeneous economy is such that:

- Household choices satisfy the recursive formulation of their problem (equations 31 and 32) given evolution of future aggregate capital, net bond receipts and current states of the world.
- Aggregate capital stock is consistent with the sum of all the household capital savings (equations 35 and 36).
- Household decisions to default or not are such that the no profit condition of the lenders is satisfied.

$$q(K_{t+1}, D_{t+1}, Z_t) = \int_{z \in def} \frac{1 - f(Z)dZ}{1 + rf_t}. \tag{42}$$

7.2 Results

In this economy, default vote tally is the sum of all the households who benefit more under default. The simplest decision is for the the age G year olds who will not survive beyond current period. Their decision is static as it only depends upon the current period consumption. In the case of no default their non bond income, which consists of rental returns and pension, is higher pension due to the presence of output loss under default. If the bond disbursement is positive i.e. capital inflows, then their income is even higher. If the economy defaults, the rental returns and pension drop due to the output loss that occurs with default. However if there are capital outflows a default would increase consumption in the amount of foregone debt service. Households of age G will not vote for default in any states in which the government provide positive bond disbursements.

Proposition 1. *Households of age G will only vote default iff* $q(K_{t+1}, D_{t+1}, Z_t)D_{t+1} - D_t < 0$.

Proof:

$$\begin{aligned}
y_t &= (1 + r_t)k_t + ss_t \\
y_t^a &= (1 + r_t^a)k_t + ss_t^a \\
y_t &> y_t^a \quad \forall t \\
B_t &= q(K_{t+1}, Z_t, D_{t+1})D_{t+1} - D_t \\
V^{R,G} &= U(y_t + B_t) < V^{D,G} = U(y_t^a) \quad iff \quad B_t < 0
\end{aligned}$$

Corollary 1. *Households of age G will only vote default iff* $(r_t^a - r_t)k_t + (ss_t^a - ss_t) < B_t$ *and the surplus from defaulting* $V^D - V^R$ *is decreasing in their capital for given* B_t*. This states that the households of age G will only vote default if the foregone debt payments are more than the capital return loss and pension reduction due to lower aggregate shock until default* $\psi < 1$

Proof:

$$\begin{aligned}
V^{R,G} &= U((1 + r_t)k_t + ss_t + B_t) \\
V^{D,G} &= U((1 + r_t^a)k_t + ss_t^a) \\
V^{D,G} &> V^{R,G} \quad iff \quad (r_t^a - r_t)k_t + (ss_t^a - ss_t) < B_t
\end{aligned}$$

Proposition 2. *Let* $y_t^a = r_t^a k_t + ss_t^a$

For the retired, default income is increasing in aggregate capital stock iff $k_t > k_t^*$ *and decreasing in ratio of retired to working*

Proof:

rental return is independent of retired and working ratio.

$$ss_t^a = \frac{\sum_{i=1}^{2}\sum_{s=1}^{R-1} mass(i,s)}{\sum_{s=R}^{G} mass(i,s)} \tau_s w_t^a n_t \tag{43}$$

$r_t^a k_t > ss_t^a$ *iff*

$$k_t > k_t^* = \frac{1-\alpha}{\alpha}\tau_s \frac{\sum_{i=1}^{2}\sum_{s=1}^{R-1} mass(i,s)}{\sum_{s=R}^{G} mass(i,s)} n_t \frac{K}{N} \tag{44}$$

$$\frac{\partial k_t^*}{\partial K_t} > 0 \tag{45}$$

Figure 10 displays the surplus $V^R - V^D$ of households of age G for some states in which they prefer default. Figure 10 shows for low capital stock, households age G will vote default if they have low wealth. The high rental returns as a result of the low aggregate capital stock will increase the cost of defaults for households with assets. As aggregate capital stock increases, defaults are preferred for higher aggregate shocks. This is due to two reasons. First high aggregate shock lowers the reduction of rental returns due to default. Under low aggregate capital stock, the rental returns are higher making defaults costlier. Second combination of high aggregate shock and aggregate capital stock increase pension enough to offset the loss of rental income.

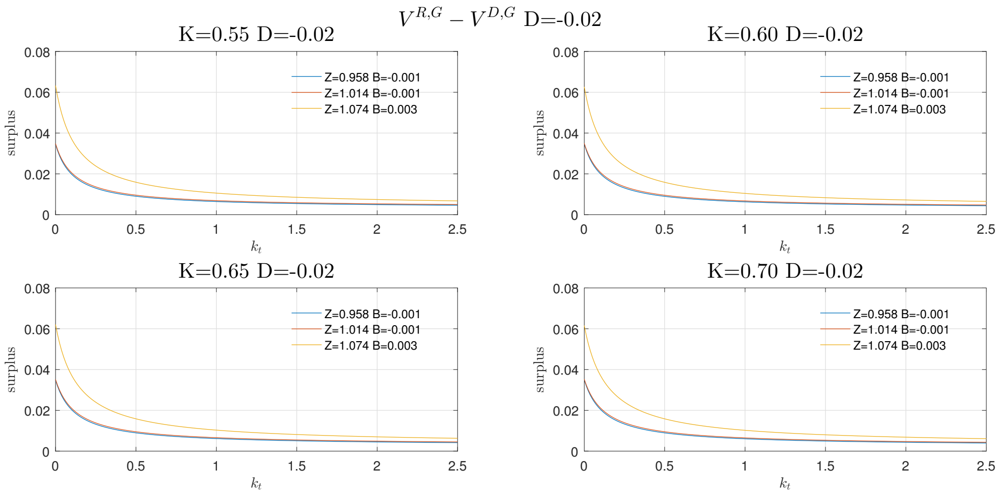

Figure 10: Surplus for Age G

In the case of the Greek referendum, increase in age implied a decrease in probability of ''No" vote. This could be reconciled with a model where production loss occurs with defaults and capital stock is sufficiently low and/or bond pricing schedule is decreasing in aggregate capital stock. A larger capital stock is induced by higher discount rate, which implies less dependence on foreign bonds due to less desire to tilt consumption forward. Less dependence on foreign bonds will make the borrower/lender relationship easier to sever.

For households that are retired of age less than G, their decision to default is dynamic. The increase in consumption current period from default will be offset by the future loss in pension. Future rental return are not necessarily decreasing. For most scenarios rental returns will be decreasing in future periods. The loss of production inhibits future capital accumulation for most periods. For sufficiently low capital stock and high technology shock, the rental rates can be higher next periods than an economy under good standing with same capital stock and technology shock but with some debt balance upper right 13. The higher rental rates can induce savings even though wages are lower. The lower wages however do not lower labor hours as there is no substitution between leisure and labor (inelastic labor supply). Figure 13 shows the comparison of aggregate capital stock for few interesting states. The aggregate capital stock under no default is perfectly horizontal for some debt. This is the result of economy being in default in those states, which implies bond prices will be zero. UR Figure 13 shows a slight increase in aggregate capital stock under default in this scenario.

For households of age $G-2$ and age R their surplus is plotted in Figures 11 and 12 respectively. For households of age R, the asset cutoff for default will be lower than that of households of age $G-2$ and age G. Less wealth is needed to favor default as they will experience lower rental returns for more periods. The age R and $G-2$ households prefer default for higher aggregate shock than the age G household. The high persistence of the stochastic process mitigates the loss of rental returns from default. The surplus from default is decreasing with age for the retired households. The reduction in rental returns in period 1 is mitigated by the higher returns following periods if aggregate capital stock is growing faster than under no default.

The retired in this economy prefer no default reconciling the correlation of lower ‘’No” with higher age in the Greek referendum. For retired households their preference to default is decreasing in wealth, which reconciles with the referendum results in the general dimension of wealth.

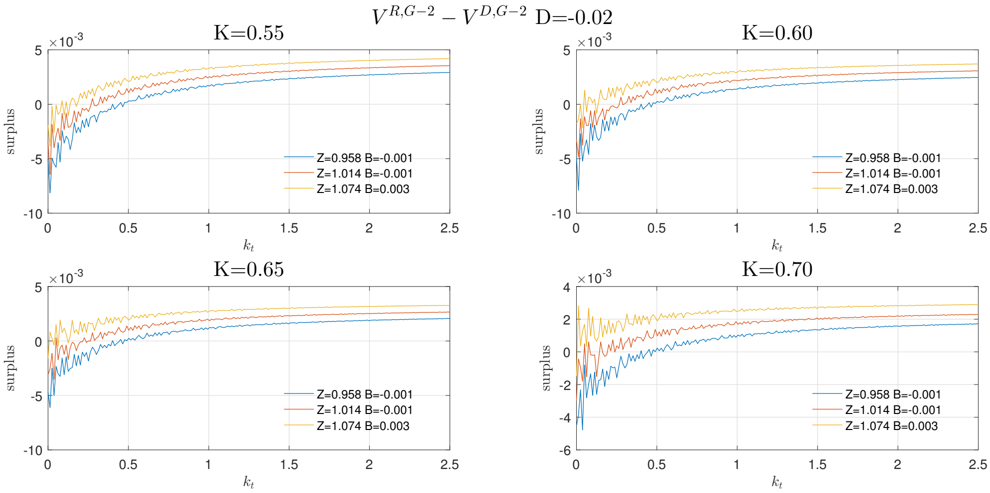

Figure 11: Surplus for Age G - 2

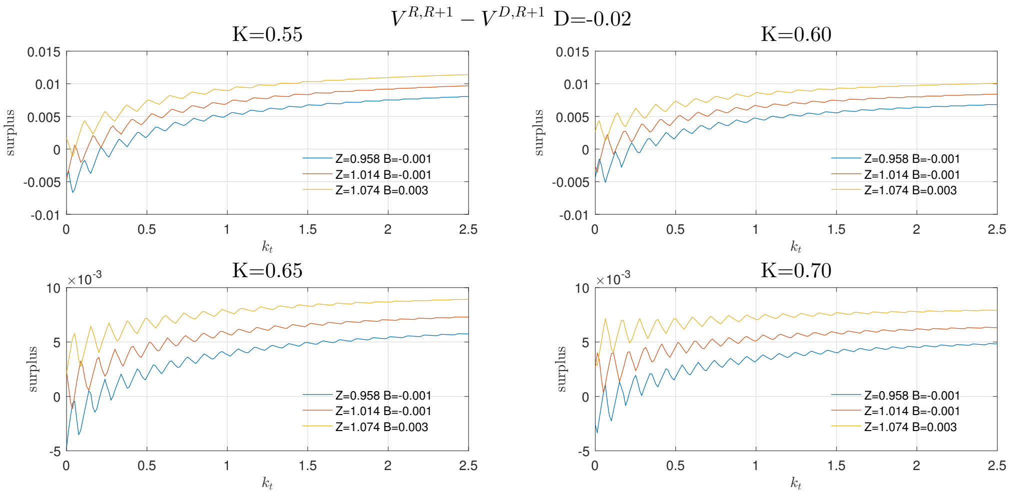

Figure 12: Surplus for Age R

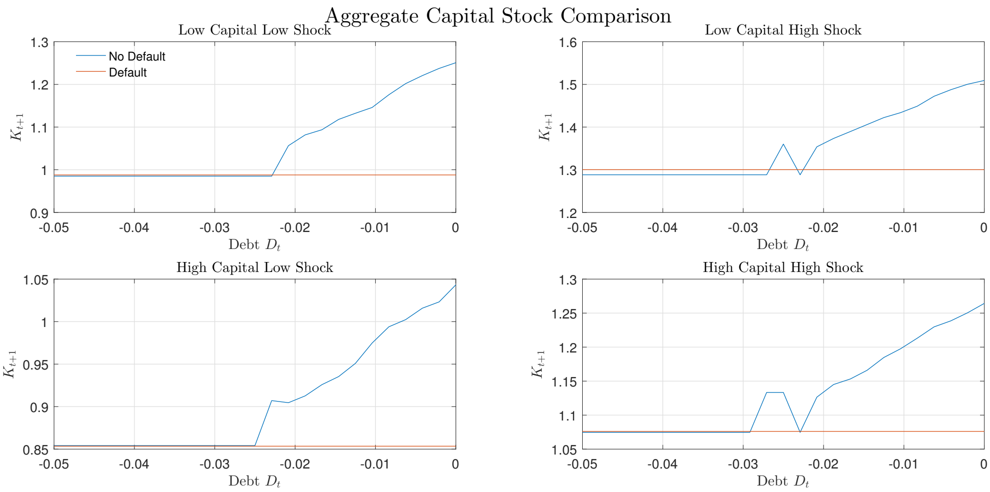

Figure 13: Evolution of Aggregate Capital

In the model economy the retirees account for less than 25% of the population. For the working, their analysis is more complicated as they have to take into account wage income, productivity type, and an idiosyncratic shock which are absent from the retirement problem. At birth, households start with zero wealth as there are no bequests in the model. For young households main source of income is wage as their rental income is low during early stages of capital accumulation. There are two types of households and two types of idiosyncratic shocks. Households that are low productive and experience low idiosyncratic shock are depicted in the UL of Figures 14, 15, and 16. These households have the largest surplus from defaults. Due to their low income, the foregoing of debt service increases marginal utility for the current period more than other types of households. As their wealth increases, the increase in consumption due to default will be less beneficial as they are trading off larger losses on their asset returns. Households that are either more productive or experience high idiosyncratic shock have smaller gain in marginal utility from the increase in current period consumption from default.

The working households prefer default more than the retired households for few reasons. First, the lower rental returns are the not as significant because young households have not accumulated much capital. Second, the lower income households gain larger marginal utility from foregone debt service. Whereas the retired are not differentiated by low and high productivity, half of the working households are low. These households have more to gain from foregone debt service. Although the working households generally favor default, note they also prefer no default with increase in wealth and income. Higher income implies a larger drop in wages due to production loss that occurs with default.

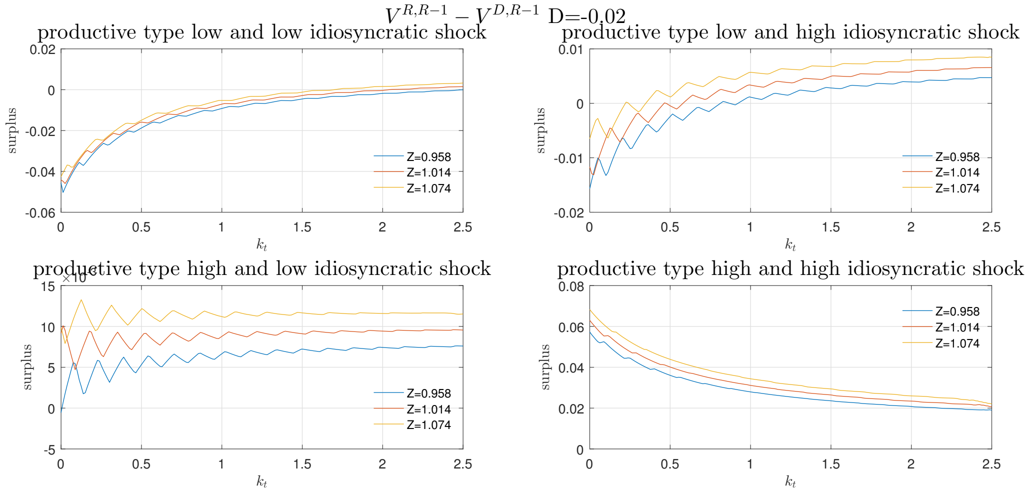

Figure 14: Surplus for Age R-1

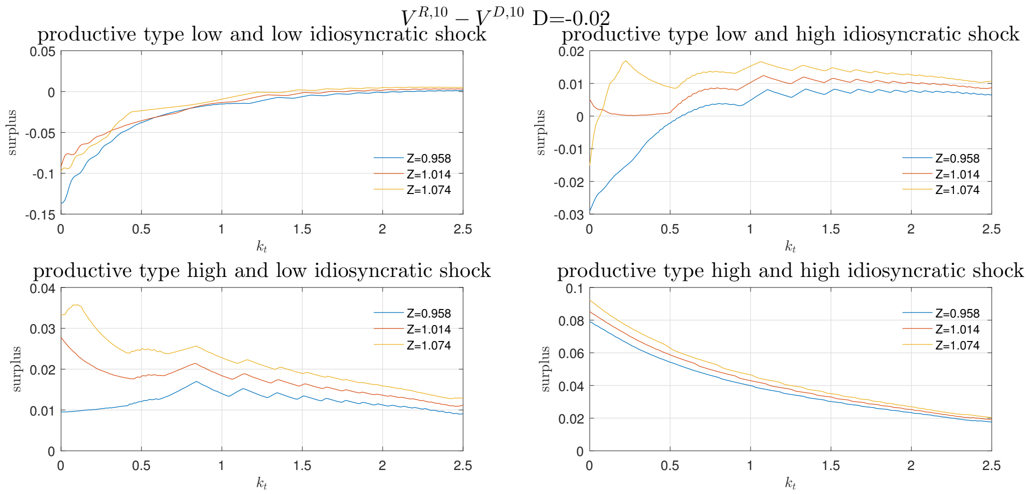

Figure 15: Surplus for Age 10

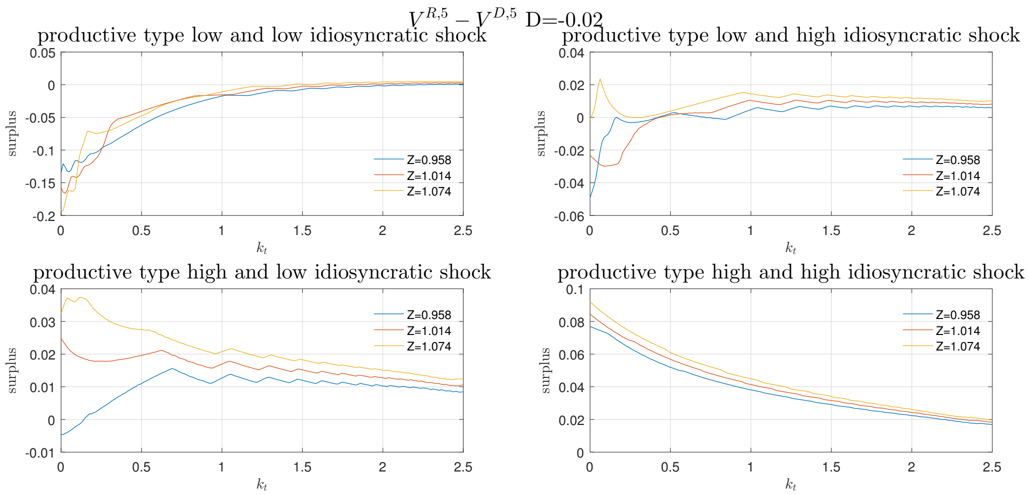

Figure 16: Surplus for Age 5

The model was unable to generate vote results for default that was similar in the range of Greek referendum $\sim 60\%$. A successful default vote was supported by more than 90% of the population. There were some many states in which the voters preferred no default by a margin of less than 10%. Following figures will explore the result of a vote favoring no default with slim margin. Figure 17 shows the proportion of the age groups that voted for default in the model. Under low capital stock and low aggregate shock, the young have a strong preference for default. The older working groups are less inclined to vote for default as they have accumulated wealth. The retirees favor no default in order to increase rental returns on their savings. Figure 18 shows the proportion of population that voted for default by wealth. There is a decreasing trend of vote share for default as wealth increases. Both figures are exhibit similar characteristics with the referendum results: vote share decreasing in age and in wealth.

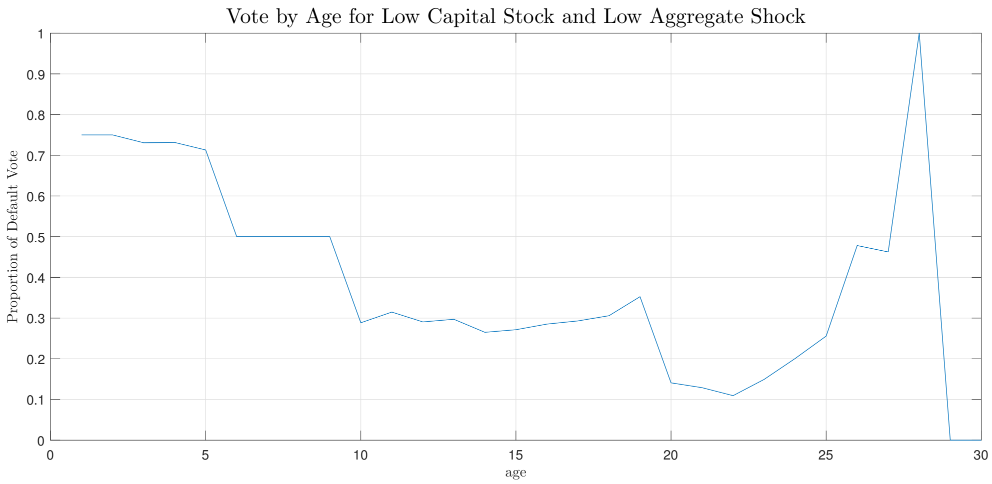

Figure 17: Vote by Age

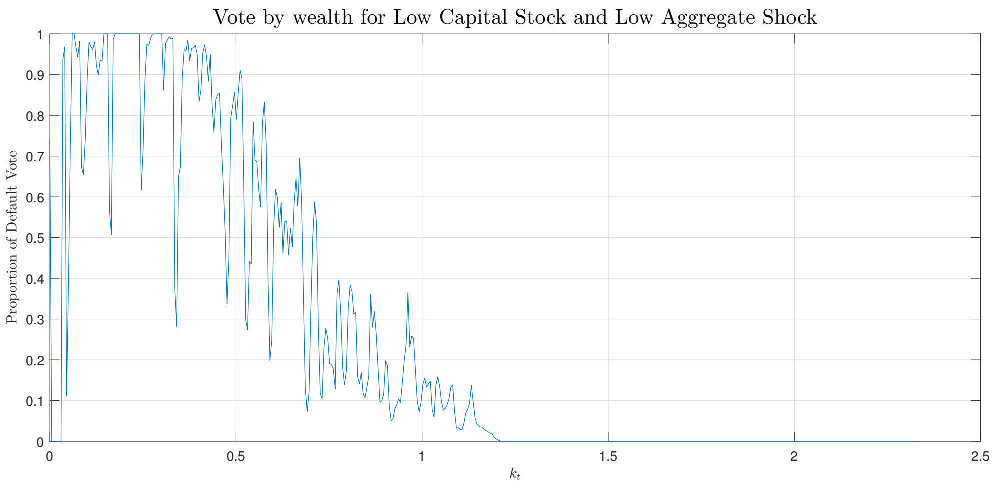

Figure 18: Vote by Wealth

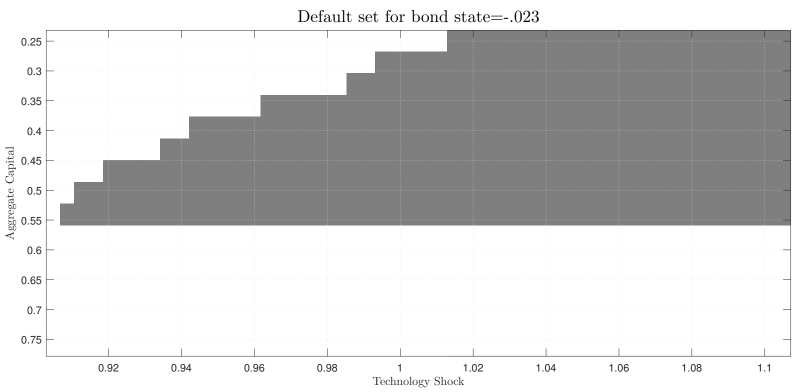

Figure 19: Default Set: Heterogeneous Economy

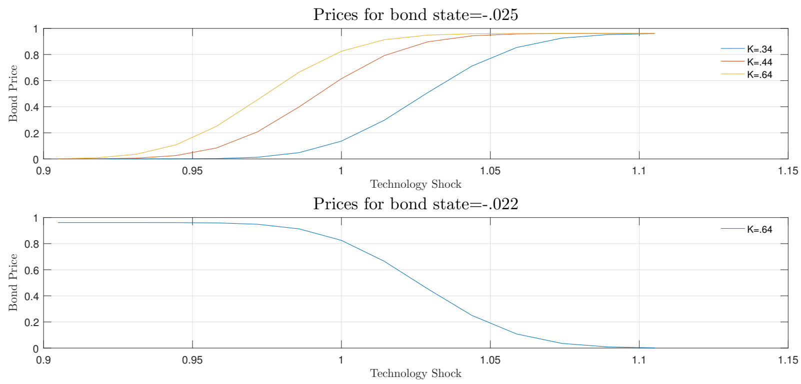

Figure 20: Bond Price: Heterogeneous Economy

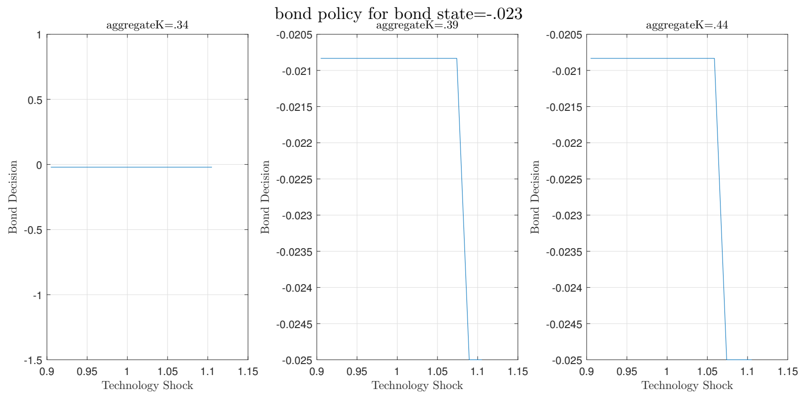

Figure 21: Bond Choice: Heterogeneous Economy

In the model the population had constant birth rate. Figure 22 shows the population density by age for the model economy. A model which exhibits a decreasing birth rate and an increase in survival probabilities a feature in most economies, the population will be older. The results the model imply an aging population is less likely to vote default as the aging population favors scenarios with higher returns on their savings.

The wealth distribution generated by the model is left skewed but does not exhibit high income inequality. The gini coefficient for the model is .4271 not too different from the actual value of .367[10]. The model implies an economy with more concentrated wealth i.e. higher proportion of the population with smaller assets is more likely to favor default.

[10]http://data.worldbank.org/indicator/SI.POV.GINI?locations=GR

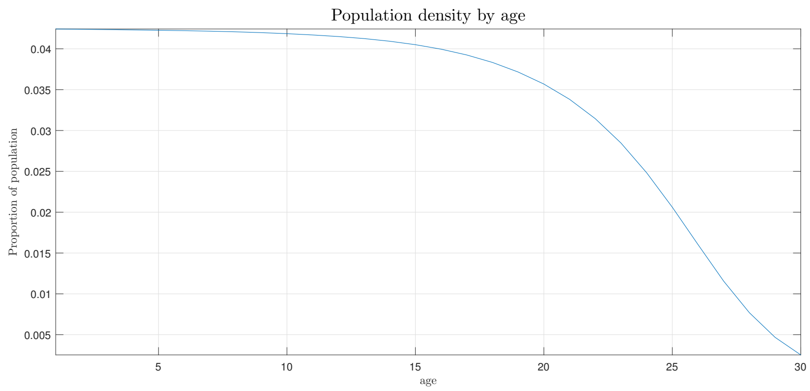

Figure 22: Population Density by Age

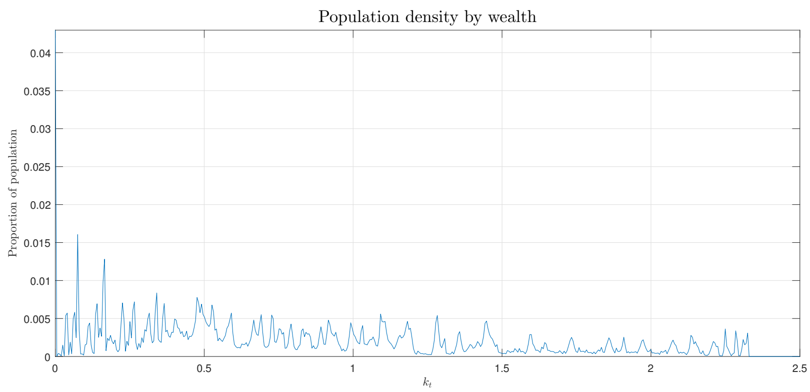

Figure 23: Population Density by Wealth

8 Conclusion

This chapter tried to model an economy with heterogeneous agents in Eaton and Gersovitz (1981) framework. By including age and mechanisms to propagate wealth distribution such as different productive types and idiosyncratic wage shocks, the distribution of wealth and age on sovereign default decision was studied. In similar fashion to the referendum results, the model showed decreasing vote share with wealth and age.

9 Appendix

The codes for the paper are hosted on https://github.com/slee126/votetodefault.

Calibration

Parameter	Description	Value
β	Discount rate	.95
δ	Depreciation rate	.1
α	capital share of output	.33
τ_w	wage tax	.3
τ_k	other income tax (pension and capital)	.3
τ_p	social security tax	.2
ρ^a	AR(1) persistence of stochastic aggregate process	.8
σ^a	standard deviation of stochastic aggregate process	.02
π^{hl}	Transition matrix for idiosyncratic process	$\begin{bmatrix} 0.9604 & .0396 \\ .0396 & 0.9604 \end{bmatrix}$
$i = l$	low productivity type wage share	.57
$i = h$	high productivity type wage share	1.43
$e_t = l$	low productivity shock wage share	.68
$e_t = h$	high productivity shock wage share	1.32
G	maximum age of household	30
R	age of retirement	20
n_t	inelastic labor provided	.33
θ	exogenous probability of re-entry to debt market	.3
ψ	production output multiplier under default	.969

Table 1: Calibrated values in the model

References

Aguiar, M. and Amador, M. (2011), "Growth in the Shadow of Expropriation," *The Quarterly Journal of Economics*, 126, 651–697.

Aguiar, M. and Gopinath, G. (2006), "Defaultable Debt, Interest Rates and Current Account," *Journal of International Economics.*

Aiyagari, R. (1995), "Optimal Capital Income Taxation with Incomplete Markets , Borrowing Constraints , and Constant Discounting," *Journal of Political Economy*, 103, 1158–1175.

Arellano, C. (2008), "Default Risk and Income Fluctuations in Economies Emerging," *American Economic Review*, 98, 690–712.

Auerbach, A. and Kotlikoff, L. (1987), "Dynamic Fiscal Policy," *Cambridge University Press.*

Borensztein, E. and Panizza, U. (2008), "The Costs of Sovereign Default," *IMF Working Paper.*

Broner, B. F., Martin, A., and Ventura, J. (2010), "Sovereign Risk and Secondary Markets," *American Economic Review*, 100, 1523–1555.

Bulow, J. and Rogoff, K. (1989), "A Constant Recontracting Model of Sovereign Debt," *Journal of Political Economy*, 97, 155.

Cuadra, G. and Sapriza, H. (2008), "Sovereign default, interest rates and political uncertainty in emerging markets," *Journal of International Economics*, 76, 78–88.

Eaton, J. and Gersovitz, M. (1981), "Debt with Potential Analysis Repudiation : Empirical Theoretical and," *The Review of Economic Studies*, 48, 289–309.

Guembel, A. and Sussman, O. (2009), "Sovereign debt without default penalties," *Review of Economic Studies*, 76, 1297–1320.

Heer, B. and Maußner, A. (2011), "The Burden of Unanticipated Inflation: Analysis of an Overlapping-Generations Model With Progressive Income Taxation and Staggered Prices," *Macroeconomic Dynamics*, 16, 278–308.

Huggett, M. (1993), "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17, 953–969.

Kletzer, K. and Wright, B. (2000), "Sovereign Debt as Intertemporal Barter," *American Economic Review*, 90.

Krusell, P. and Smith, A. (1998), "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

Levy-yeyati, E. and Panizza, U. (2006), "The Elusive Costs of Sovereign Defaults," .

Livshits, I., Phan, T., and Trebesch, C. (2014), "Sovereign Default and Political Turnover (preliminary and incomplete)," .

McKelvey, R. (1979), "General Conditions for a Global Intransitivities in Formal Voting Models," *Econometrica*, 47.

Tabellini, G. (1991), "The Politics of Intergenerational Redistribution ," *Journal of Political Economy*, 99, 335–357.

Tomz, M. and Wright, M. L. J. (2007), "Do Countries Default in Bad Times? ," *Journal of the European Economic Association*, 5, 352–360.

Yue, V. Z. (2010), "Sovereign default and debt renegotiation," *Journal of International Economics*, 80, 176–187.

Zettelmeyer, J., Trebesch, C., and Gulati, M. (2013), "The Greek debt restructuring: An autopsy," *Economic Policy*, 28, 513–563.

Chapter 2: Relative Tick

This chapter is a joint paper with Professor Eric Aldrich of University of California Santa Cruz. The unpublished article is inserted with his permission.

Abstract

We develop a theoretical model to highlight a previously unexplored mechanism of price discovery: relative minimum price increments for equivalent assets trading on distinct financial exchanges. Although conventional wisdom dictates that futures market assets lead equities equivalents in terms of price formation, our model predicts that the opposite should be true when particular relative price conditions hold for the bids and offers of each asset. We develop a new empirical measure of price discovery which is suited to asynchronous, high-frequency transaction and quotation data, and apply it to the highly liquid E-mini/SPY pair in order to test the predictions of the model. Empirical evidence strongly supports the model and further demonstrates that relative minimum contract size plays an additional role in the formation of prices.

Keywords: Market microstructure, market design, high-frequency trading, entropy.

JEL Classification: G12, G14.

10 Introduction

On May 10, 2010, four days after the Flash Crash, the CME Group issued a staff report explaining precautions taken at the Chicago Mercantile Exchange (CME) during the volatile events of May 6, 2010. According to the staff report, "The primary purposes of futures markets are to provide an efficient mechanism for price discovery and risk management" and "...stock index futures frequently represent the venue in which price information is revealed first, generally followed closely by spot markets" (Labuszewski and Co, 2010). Indeed, conventional trading wisdom dictates that "futures markets lead cash markets".

Academic work has largely supported the statements of the CME Group staff report. Kawaller et al. (1987) and Stoll and Whaley (1990) are early examples, demonstrating a leading relationship between S&P 500 index futures and the S&P 500 index itself. Hasbrouck (2003) presents similar findings for small-denomination S&P 500 index futures contracts (E-minis) and equities market exchange traded funds (ETF) that track the S&P 500 index. Futures and cash markets for Canadian bonds are shown by Campbell and Hendry (2007) to behave in like manner. More recently, Laughlin et al. (2014) and Aldrich et al. (2016) use econometric methodology similar to that of this paper to articulate the tightly coupled relationship of messaging traffic between futures and equities exchanges.

While the "futures-leads-cash" relationship is widely considered a market standard, counterexamples exists. Stephan and Whaley (1990), Easley et al. (1998) and Chakravarty et al. (2004) document a reversal of informed trading for single stock options: equities lead the derivatives market. Easley et al. (1998), however, show that this reversal is an unconditional fact and is not true under specific put/call trade conditions. Yang (2009) shows the same reversal of informed trading to be true of currency markets.

The primary objective of this work is to contribute to current knowledge on the determinants of price discovery. In particular, we seek to understand why price discovery often occurs primarily in a single venue, despite the fact that a single asset, or its equivalents, trade in diverse venues. A standard explanation for the primacy of derivatives markets for price discovery is the ability of informed traders to exploit greater leverage. Fleming et al. (1996) argues that price discovery should occur in the market with lower trading costs as measured by bid/offer spread, broker commissions and market impact of large orders. Following this logic, a wider minimum price increment (higher cost) would be associated with diminished price discovery.

Although contract specifications are uniform for stocks that trade on diverse exchanges in the National Market System (NMS), the same is not true across derivatives and cash (equities or currency spot) markets. After a simple basis adjustment (to account for dividends and interest), many pairs of futures and cash instruments may be considered identical, despite the fact that they trade in distinct locations. Further, because futures and cash instruments are regulated by different entities, they frequently differ with respect to contract specifications, such as minimum price increment, minimum contract size, notional value, etc.

A primary contribution of this paper is a model that highlights the relationship of relative minimum price increments to price discovery. Contrary to the reasoning above, that larger price increments should be associated with diminished information share, our model suggests a countervailing effect *when the minimum increment in one market is larger than that of another.* The mechanism relies on simple deterministic arbitrage among simultaneously posted bids and offers in each market. According to these arbitrage arguments, the model makes several specific predictions about the direction of informed trade and under what conditions price discovery reversals should occur.

Our second contribution is to refine the econometric methodology for detecting lead/lag behavior between financial exchanges. The current standard for measuring price discovery is developed in Hasbrouck (1995), which expresses fragmented market prices as a cointegrated system and which formally interprets market information share as the fraction of price variation attributed to the permanent innovation in each market. While this methodology is intuitively appealing when price discovery is viewed through the lens of variance contribution, it has a number of weaknesses. First, it requires synchronization of price observations across the cointegrated series.

This is typically done by measuring prices in clock time. As demonstrated by Ane and Geman (2000) and Aldrich et al. (2015), transaction-time models are typically superior to clock-time models, especially for fine-grained, high-frequency applications. Second, VAR methodology imposes a parametric model. Third, identification issues often arise when incorporating sufficient lag information in the system. Hasbrouck (1995) deals with this by imposing polynomial lag restrictions on the system coefficients. Finally, when adapting the method to recent, high-frequency data, the identification and estimation issues are compounded.

We introduce a method that is model free and well suited for high-frequency data. While the computational burden is not trivial, it does not suffer the identification issues that arise in richly parameterized vector autoregression systems. Further, it is not narrowly interpreted as a measure of variance decomposition, but is a direct measure of lead/lag relationships between transactions at distinct trading centers. It is specifically designed to deal with asynchronous data and implicitly tests a null hypothesis of no leading information in the transactions of a particular market.

Our final contribution is to apply our econometric methodology to test the model predictions on the liquid E-mini S&P 500 index futures/SPDR ETF pair (tickers ES and SPY, respectively). Our empirical work is especially careful with the issue of simultaneity, accounting for speed-of-light transmission latency between the CME Globex matching engine in Aurora Illinois, and the equities exchanges in New Jersey (we focus on Nasdaq OMX in Carteret). We find strong evidence in favor of the model, suggesting that relative minimum increments are an extremely important factor in terms of determining the share of informed trading and price discovery within a market. We further highlight the secondary effect of relative contract size, which accentuates/attenuates the share of price discovery.

Our results are supportive of recent work by Hagströmer and Menkveld (2016), who extend the methods of Hasbrouck (1995) to build a network map of information percolation, estimating not only the relative weights of market centers and participants in the FX market, but also the relative flow of information between them. While our approach is somewhat distinct, we likewise find that intermediaries are a fundamental component of price cointegration. Specifically, although price cointegration among fragmented markets is the result of a variety of factors, arbitrage opportunities within an information network of prices causes intermediaries to be a critical vehicle for maintaining price uniformity.

Fundamentally, rather than questioning how markets are organized and how information flows between participants and exchanges, our work questions why those flows are observed. Most importantly, we view our results as being of primary relevance to market regulators. In October 2016, the Financial Industry Regulatory Authority (FINRA) will implement the Tick Size Pilot Program, as ordered by the Securities and Exchange Commission (SEC). Under the program, several test groups of small capitalization equities will be required to quote, and potentially transact, at wide price increments of \$0.05, rather than the standard \$0.01. Understanding the role of such price increments both within and across markets is important for regulators such as the SEC and Commodity Futures Trading Commission (CFTC), for whom policy coordination may be necessary in order to promote stable and well-functioning markets.

11 Model

11.1 Environment

The environment is comprised of a single asset, S, and a futures contract on the asset, F. The asset and futures contract trade in distinct markets separated by communication latency τ. The markets are populated by three agents, who are distinguished by the following types: a market maker for F, a market maker for S and an investor, who takes the role of an informed trader. All agents maximize linear utility

$$\mathcal{U}(\mu, \sigma) = \mu - \frac{1}{2}\gamma\sigma^2. \tag{46}$$

In each market, $X \in \{F, S\}$, at time t, the market maker offers $q_{m,t}^{X,o}$ units of the asset for sale at price $p_{m,t}^{X,o}$ and bids to purchase $q_{m,t}^{X,b}$ at price $p_{m,t}^{X,b}$. She assesses the fair value of the asset to be $p_{m,t}^{X,f}$. The investor demands $q_{i,t}^{X,o}$ units of the asset at the market maker offer price and supplies $q_{i,t}^{X,b}$ units at the market maker bid. The difference between bid and offer prices is known as the spread and is denoted $2\xi_t^X$ (i.e. ξ_t^X is the half spread). We make the following assumptions:

Assumption 11.1. *Market maker assessment of the fair price in market $X \in \{F, S\}$ is an increasing function of investor demand to buy in that market, $\partial p_{m,t}^{X,f} / \partial q_{i,t}^{X,o} > 0$, and a decreasing function of investor supply to sell in that market, $\partial p_{m,t}^{X,f} / \partial q_{i,t}^{X,b} < 0$.*

Assumption 11.2. *Investor demand is a function of the bid/offer spread, with quantities demanded at the bid and offer decreasing with the size of the spread: $\partial q_{i,t}^{X,o} / \partial \xi_t^X < 0$ and $\partial q_{i,t}^{X,b} / \partial \xi_t^X < 0$ for $X \in \{F, S\}$.*

Assumption 1 is a reflection of adverse selection: uninformed market makers adjust their assessment of fair market valuation with informed investor order flow. Assumption 2 is a reflection of trading costs: as costs increase, the informed investor demands less. The following proposition will allow us to simplify notation:

Proposition 11.3. *The fair price for the market maker is equidistant between the posted bid and offer in each market: $p_{m,t}^{X,f} = p_{m,t}^{X,o} - \xi_t^X = p_{m,t}^{X,b} + \xi_t^X$, $X \in \{F, S\}$.*

Proof. By symmetry of the market maker problem outlined in Section 11.3, and the fact that they are uninformed about the direction of investor order flow, market makers achieve optimality by symmetrically placing bids and offers around their fair valuation of the assets, F and S. □

To ease notation in the sequel, we will only outline optimization problems for transactions at the best offer: a purchase by an investor and a sale by a market maker. The analogous problems for transactions on the bid are symmetric. We thus reduce notation in the following manner: $q_{m,t}^X = q_{m,t}^{X,o}$, $q_{i,t}^X = q_{i,t}^{X,o}$, and $p_{m,t}^X = p_{m,t}^{X,f}$, for $X \in \{F, S\}$.

11.2 Investor

The investor earns returns in market $X \in \{F, S\}$ by purchasing the asset at the current offer price, $p_{m,t}^X + \xi_t^X$, and liquidating at the subsequent bid price, $p_{m,t+1}^X -$

ξ^X_{t+1}. We denote the realized return and its first two moments as

$$r^X_{i,t+1} = p^X_{m,t+1} - \xi^X_{t+1} - p^X_{m,t} - \xi^X_t \tag{47}$$

$$\mu^X_{i,t} = \mathbb{E}_i\left[p^X_{m,t+1} - \xi^X_{t+1}\right] - p^X_{m,t} - \xi^X_t \tag{48}$$

$$\sigma^{2,X}_{i,t} = \text{Var}_i\left(r^X_{i,t+1}\right). \tag{49}$$

Holding period returns over longer horizons are simply the sum of single-period realized returns. We make the assumption that assets F and S are perfectly correlated under the investor's subjective assessment:

Assumption 11.4. $Cov_i\left(r^F_{i,t+1}, r^S_{i,t+1}\right) = \sigma^F_{i,t}\sigma^S_{i,t}$.

Given a budget, B, the investor allocates resources through share purchases, $q^F_{i,t}$, $q^S_{i,t}$:

$$\max_{q^F_{i,t}, q^S_{i,t}} \mu_{i,t} - \frac{1}{2}\gamma\sigma^2_{i,t} \tag{50a}$$

subject to

$$\mu_{i,t} = \left(\mathbb{E}_i\left[p^F_{m,t+1} - \xi^F_{t+1}\right] - p^F_{m,t} - \xi^F_t\right) q^F_{i,t} + \left(\mathbb{E}_i\left[p^S_{m,t+1} - \xi^S_{t+1}\right] - p^S_{m,t} - \xi^S_t\right) q^S_{i,t} \tag{50b}$$

$$\sigma^2_{i,t} = \left(q^F_{i,t}\sigma^F_{i,t}\right)^2 + \left(q^S_{i,t}\sigma^S_{i,t}\right)^2 + 2q^F_{i,t}q^S_{i,t}\sigma^F_{i,t}\sigma^S_{i,t}. \tag{50c}$$

$$q^F_{i,t}\left(p^F_{m,t} + \xi^F_t\right) + q^S_{i,t}\left(p^S_{m,t} + \xi^S_t\right) \le B. \tag{50d}$$

Equations (50b) and (50c) are the expected return and variance of the portfolio. In addition to the budget constraint, (50d), bounding inequality constraints exist for the control variables, $q^F_{i,t}$ and $q^S_{i,t}$; we have intentionally ignored these constraints as they enter the first-order conditions as constants and add little value to the economic digression of the model. The first-order conditions are

$$\mathbb{E}_i\left[p^F_{m,t+1} - \xi^F_{t+1}\right] - p^F_{m,t} - \xi^F_t - (1+\lambda^B_{i,t})\frac{\partial p^F_{m,t}}{\partial q^F_{i,t}} q^F_{i,t} \\ - \gamma\sigma^F_{i,t}\left(q^F_{i,t}\sigma^F_{i,t} + q^S_{i,t}\sigma^S_{i,t}\right) - \lambda^B_{i,t}(p^F_{m,t} + \xi^F_t) = 0 \tag{51}$$

$$\mathbb{E}_i\left[p^S_{m,t+1} - \xi^S_{t+1}\right] - p^S_{m,t} - \xi^S_t - (1+\lambda^B_{i,t})\frac{\partial p^F_{m,t}}{\partial q^S_{i,t}} q^S_{i,t} \\ - \gamma\sigma^S_{i,t}\left(q^S_{i,t}\sigma^S_{i,t} + q^F_{i,t}\sigma^F_{i,t}\right) - \lambda^B_{i,t}(p^S_{m,t} + \xi^S_t) = 0, \tag{52}$$

where $\lambda^B_{i,t}$ is the Lagrange multiplier for constraint (50d). Equations (51) and (52) result in the following solutions:

$$q^F_{i,t} = \Big(\mathbb{E}_i\left[p^F_{m,t+1} - \xi^F_{t+1}\right] - p^F_{m,t} - \xi^F_t \\ - \gamma q^S_{i,t}\sigma^F_{i,t}\sigma^S_{i,t} - \lambda^B_{i,t}(p^F_{m,t} + \xi^F_t)\Big) \times \left(\gamma\sigma^{2,F}_{i,t} + (1+\lambda^B_{i,t})\frac{\partial p^F_{m,t}}{\partial q^F_{i,t}}\right)^{-1} \tag{53}$$

$$q^S_{i,t} = \Big(\mathbb{E}_i\left[p^S_{m,t+1} - \xi^S_{t+1}\right] - p^S_{m,t} - \xi^S_t \\ - \gamma q^F_{i,t}\sigma^S_{i,t}\sigma^F_{i,t} - \lambda^B_{i,t}(p^S_{m,t} + \xi^S_t)\Big) \times \left(\gamma\sigma^{2,S}_{i,t} + (1+\lambda^B_{i,t})\frac{\partial p^S_{m,t}}{\partial q^S_{i,t}}\right)^{-1}. \tag{54}$$

Equations (53) and (54) show that investor demand increases with expected return in the respective assets, decreases in the volatility of both assets (both own and cross volatility), decreases with the elasticity of price to investor demand ($\partial p^X_{m,t}/\partial q^X_{i,t}$), and decreases with demand of the other asset.

11.3 Market Maker

Market makers are required to continually post bids and offers in the F and S markets. They are compensated for their services via the bid/offer spreads, ξ^F and ξ^S, which are their control variables. At the time of a passive sale in market $X \in \{F, S\}$ at the current offer price, $p^X_{m,t} + \xi^X_t$, the market maker earns returns via two mechanisms: (1) retaining some fraction of the sale quantity for a passive purchase (with an aggressive investor) in her own market in the subsequent period, and (2) aggressively purchasing some fraction of the sale quantity at the offer price in the other market, X^c, at time $t+\tau$, $p^{X^c}_{m,t+\tau} + \xi^{X^c}_{t+\tau}$. The time shift τ accounts for communication latency between the two markets. Since the market maker's own-market transactions are passive, the expected repurchase price is her expected fair valuation, which is her long-term, passive transaction price. This is equivalent to saying that with equal probability, she will transact at the bid, $\mathbb{E}_m\left[p^X_{m,t+1} - \xi^X_{t+1}\right]$, and offer, $\mathbb{E}_m\left[p^X_{m,t+1} + \xi^X_{t+1}\right]$, in the subsequent period, which nets out to an expected repurchase price of $\mathbb{E}_m\left[p^X_{m,t+1}\right]$. We make the following assumption.

Assumption 11.5. *Under the subjective assessment of the market maker, the fair price of each asset is a martingale:* $\mathbb{E}_m\left[p^X_{m,t+1}\right] = p^X_{m,t}$, $X \in \{F, S\}$.

As uninformed market participants, Assumption 11.5 is natural for the market makers. We denote the realized returns and their first two moments as

$$r^X_{m,t+1} = p^X_{m,t} + \xi^X_t - p^X_{m,t+1} \tag{55a}$$

$$\mu^X_{m,t+1} = p^X_{m,t} + \xi^X_t - \mathbb{E}_m\left[p^X_{m,t+1}\right] = \xi^X_t \tag{55b}$$

$$\sigma^{2,X}_{m,t+1} = \mathrm{Var}_m\left(r^X_{m,t+1}\right) \tag{55c}$$

$$r^{X^c}_{m,t+\tau} = p^X_{m,t} + \xi^X_t - p^{X^c}_{m,t+\tau} - \xi^{X^c}_{t+\tau} \tag{55d}$$

$$\mu^{X^c}_{m,t+\tau} = \mathbb{E}_m\left[r^{X^c}_{m,t+\tau}\right] = p^X_{m,t} + \xi^X_t - \mathbb{E}_m\left[p^{X^c}_{m,t+\tau} + \xi^{X^c}_{t+\tau}\right] \tag{55e}$$

$$\sigma^{2,X^c}_{m,t+\tau} = \mathrm{Var}_m\left(r^{X^c}_{m,t+\tau}\right). \tag{55f}$$

Conditional on a passive sale in each market, makers oversee a control problem, which involves the choice of bid/offer spread and a portfolio of repurchases. As the two problems are symmetric, we focus on the case of a sale in the market for F. At the moment of a sale, the market maker is filled at an exogenously determined quantity, $q^F_{i,t}$, chosen by the investor. In addition to her choice of bid/offer spread, ξ^F_t, she also chooses an optimal portfolio of repurchases via the variable $q^S_{m,t+\tau}$. The resulting optimality problem is:

$$\max_{q^S_{m,t+\tau},\xi^F_t} \mu_{m,t+1,t+\tau} - \frac{1}{2}\gamma\sigma^2_{m,t+1,t+\tau} \tag{56a}$$

subject to

$$\mu_{m,t+1,t+\tau} = \xi^F_t \left(q^F_{i,t} - q^S_{m,t+\tau}\right) + \left(p^F_{m,t} + \xi^F_t - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right]\right) q^S_{m,t+\tau} \tag{56b}$$

$$\begin{aligned}\sigma^2_{m,t,t+\tau} = &\left(\left(q^F_{i,t} - q^S_{m,t+\tau}\right)\sigma^F_{m,t+1}\right)^2 + \left(q^S_{m,t+\tau}\sigma^S_{m,t+\tau}\right)^2 \\ &+2\left(q^F_{i,t} - q^S_{m,t+\tau}\right) q^S_{m,t+\tau}\sigma^F_{m,t+1}\sigma^S_{m,t+\tau}\end{aligned} \tag{56c}$$

Equation (56b) says after a passive fill at the F offer, the market maker lays off her risk by retaining $q^F_{i,t} - q^S_{m,t}$ shares of F in her own market at time $t+1$ and purchasing $q^S_{m,t}$ shares of S at time $t+\tau$. Her motive for laying off risk in the S market is that the anticipated cross-market return, $p^F_{m,t} + \xi^F_t - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right]$, may be larger than the anticipated within-market return, ξ^F_t. The cross-market hedge can be thought of as latency arbitrage profits earned by high-frequency market makers. In addition to the spreads they choose in the separate markets, it is part of their total compensation package.

The first-order conditions of the market maker's problem are:

$$\begin{aligned}p^F_{m,t} - \mathbb{E}_m&\left[p^S_{m,t+\tau} + \xi^S_{t+\tau} + \frac{\partial p^S_{m,t+\tau}}{\partial q^S_{m,t+\tau}} q^S_{m,t+\tau}\right] + \gamma\left(q^F_{i,t} - q^S_{m,t+\tau}\right)\sigma^{2,F}_{m,t+1} \\ &- \gamma q^S_{m,t+\tau}\sigma^{2,S}_{m,t+\tau} - \gamma\left(q^F_{i,t} - 2q^S_{m,t+\tau}\right)\sigma^F_{m,t+1}\sigma^S_{m,t+\tau} = 0\end{aligned} \tag{57}$$

$$\begin{aligned}q^F_{i,t} + \xi^F_t \frac{\partial q^F_{i,t}}{\partial \xi^F_t} - \gamma&\left(q^F_{i,t} - q^S_{m,t+\tau}\right)\sigma^{2,F}_{m,t+1}\frac{\partial q^F_{i,t}}{\partial \xi^F_t} \\ &- \gamma q^S_{m,t+\tau}\sigma^F_{m,t+1}\sigma^S_{m,t+\tau}\frac{\partial q^F_{i,t}}{\partial \xi^F_t} = 0,\end{aligned} \tag{58}$$

where we have allowed the S maker's fair price at $t+\tau$, $p^S_{m,t+\tau}$, to be sensitive to the F maker's order flow, $q^S_{m,t+\tau}$, since the F maker takes the role of an aggressive investor. Solving for the control variables in Equations (57) and (58):

$$q^S_{m,t+\tau} = \Big(p^F_{m,t} - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right] + \gamma q^F_{i,t}\sigma^F_{m,t+1}\left(\sigma^F_{m,t+1} - \sigma^S_{m,t+\tau}\right)\Big)D^{-1} \tag{59}$$

$$\xi^F_t = \gamma q^F_{i,t}\sigma^{2,F}_{m,t+1} - \gamma q^S_{m,t+\tau}\left(\sigma^{2,F}_{m,t+1} - \sigma^F_{m,t+1}\sigma^S_{m,t+\tau}\right) - q^F_{i,t}\left(\frac{\partial q^F_{i,t}}{\partial \xi^F_t}\right)^{-1} \tag{60}$$

where

$$D = \gamma\left(\sigma^F_{m,t+1} - \sigma^S_{m,t+\tau}\right)^2 + \frac{\partial p^S_{m,t+\tau}}{\partial q^S_{m,t+\tau}}. \tag{61}$$

Substituting Equation (59) into (60):

$$\xi^F_t = q^F_{i,t}\left(\gamma\sigma^{2,F}_{m,t+1} - \gamma^2\sigma^{2,F}_{m,t+1}\left(\sigma^F_{m,t+1} - \sigma^S_{m,t+\tau}\right)^2 D^{-1} - \left(\frac{\partial q^F_{i,t}}{\partial \xi^F_t}\right)^{-1}\right)$$
$$- \gamma\left(p^F_{m,t} - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right]\right)\sigma^{,F}_{m,t+1}\left(\sigma^F_{m,t+1} - \sigma^S_{m,t+\tau}\right)D^{-1}. \tag{62}$$

Equation (59) shows that the F market maker's optimal cross-market arbitrage quantity is determined by three primary components: (1) size of the expected arbitrage profit, (2) S market price elasticity to the cross-market arbitrage order size and (3) the relative volatilities in the two markets. The arbitrage profit is captured by the first term, $p^F_{m,t} - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right]$, which represents the difference between the expected repurchase prices in the two markets. The price elasticity, which according to Assumption 11.1 is positive, enters in the denominator, and highlights how increasing market impact in the S market (high elasticity) reduces the incentive of the F market maker to route orders to that market. The relative market volatilities have several interesting effects. First, note from the profit condition that if $\sigma^F_{m,t+1} = \sigma^S_{m,t+\tau}$, the arbitrage quantity, $q^S_{m,t+\tau}$, is only positive if the expected S offer price at $t+\tau$ is below the expected F mid price at t. Further, the only term in the denominator, which influences the size, not the sign, of the order, is the price elasticity. When $\sigma^F_{m,t+1} > \sigma^S_{m,t+\tau}$, however, the second term of Equation (59) has a positive effect, which weakens the profit requirements of the F market maker: the arbitrage order quantity can be positive even when the expected S offer price is above the expected F mid price. At the same time, the denominator increases because of the additional positive term related to volatility. The net effect on $q^S_{m,t+\tau}$ may be positive or negative. The higher relative volatility in the F market increases the potential upside gains of expected arbitrage profits, but also increases the potential downside losses. Clearly, risk aversion, captured by γ, plays an important role in determining the net effect. When $\sigma^F_{m,t+1} < \sigma^S_{m,t+\tau}$, the effect on $q^S_{m,t+\tau}$ (relative to the equal volatility case) is unambiguously negative: the second term in Equation (59) is negative, which effectively states that the F maker demands a more strict profit condition, and the first term of the denominator also increases. This latter effect is natural, as the increased S volatility, and the maker's risk aversion to it, decrease the utility from potential arbitrage gains. Additionally, we note that $q^S_{m,t+\tau}$ is increasing in investor order flow to the F market maker, $q^F_{i,t}$, and that when $\sigma^F_{m,t+1} = \sigma^S_{m,t+\tau}$, $q^S_{m,t+\tau} \to \infty$ as $\partial p^S_{m,t+\tau}/\partial q^S_{m,t+\tau} \to 0$. The latter statement highlights the fact that when S market impact of the F maker's aggressive order flow diminishes to zero, her desire to route arbitrage orders grows unboundedly. In practice, since $q^S_{m,t+\tau} \leq q^F_{i,t}$, there is a small enough price elasticity to cause the market maker's order flow to achieve the constraint.

Equation (62) highlights the mechanisms that determine the market maker's optimal spread. First, the term multiplying $q^F_{i,t}$ is always positive, since the spread elasticity of investor demand is negative, according to Assumption 11.2. The result is that investor demand induces the market maker to increase the F spread, all else equal. The magnitude of the effect is positively related to $\partial p^S_{m,t+\tau}/\partial q^S_{m,t+\tau}$ (through the denominator D), and negatively related to $\partial q^F_{i,t}/\partial \xi^F_t$. Intuitively, the market maker's spread response to increased investor demand increases with her market impact in the S and decreases with investor sensitivity to spread. The second term of the equation shows that the relationship of ξ^F_t to $\xi^S_{t+\tau}$ depends on the relative market volatilities: it is positive when $\sigma^F_{m,t+1} > \sigma^S_{m,t+\tau}$, and negative when $\sigma^F_{m,t+1} < \sigma^S_{m,t+\tau}$. This captures the market maker's desire to both increase arbitrage profits (the positive relationship) while also reducing portfolio volatility (the potentially negative relationship). As with the sensitivity to investor demand, the magnitude of the spread relationship is determined by the risk aversion parameter and the relative market volatilities.

To reconcile the empirical relationship of spreads in the F and S markets, we make the following assumption and proposition.

Assumption 11.6. *The spread elasticity of investor F demand is smaller (in magnitude) than that of S demand:* $\left|\partial q_{i,t}^F / \partial \xi_t^F\right| < \left|\partial q_{i,t}^S / \partial \xi_t^S\right|$.

Proposition 11.7. *Conditional on constant and equal market volatilities*

$$\sigma_{m,t}^F = \sigma_{m,t+1}^F = \sigma_{m,t+\tau}^S = \sigma_{m,t}^S, \tag{63}$$

and conditional on equal investor order flow arriving at each market, $q_{i,t}^F = q_{i,t}^S$*, the spread for F is greater than that of S:* $\xi_t^F > \xi_t^S$.

Proof. We denote the common, constant market volatility as $\sigma_{m,t}$ and the common investor flow as $q_{i,t}$. According to Equation (60), and exploiting the symmetry of the S market maker control problem, the equilibrium spreads are

$$\xi_t^F = \gamma q_{i,t} \sigma_{m,t}^2 - q_{i,t} \left(\frac{\partial q_{i,t}}{\partial \xi_t^F} \right)^{-1} \tag{64}$$

$$\xi_t^S = \gamma q_{i,t} \sigma_{m,t}^2 - q_{i,t} \left(\frac{\partial q_{i,t}}{\partial \xi_t^S} \right)^{-1}. \tag{65}$$

The result directly follows from Assumption 11.6. □

Assumption 11.6 states that investor order flow in the F market is less sensitive to changes in the bid/offer spread than investor order flow in the S market. Such an assumption is reasonable from a practical perspective, as it captures incentives, such as increased leverage, for investors to direct informed flow to the futures market, which we have not included in the model. The result is that under a reasonable assumption of equal or similar market volatilities in the two assets, the equilibrium spread for F will be larger than that of S. This is observed in practice (see Section 13.1).

Combining Assumption 11.6 with the preceding model results, we see that the relationship of equilibrium market spreads is determined by three primary mechanisms:

1. Expected arbitrage profits (positive effect).

2. A desire to attract more investor order flow (negative effect, through the quantity elasticity).

3. Non-modeled effects, such as leverage ($\xi^F > \xi^S$).

We conclude by summarizing the important implications of our model in a proposition. The proposition is stated only for the case of equal market volatilities, but a similar version holds when volatilities are not equal, and the market maker assessment of cross-market arbitrage profits includes a risk adjustment (the second term of Equation (59)).

Proposition 11.8. *Conditional on constant and equal market volatilities*

$$\sigma^F_{m,t} = \sigma^F_{m,t+1} = \sigma^S_{m,t+\tau} = \sigma^S_{m,t}, \tag{66}$$

the following statements hold:

1. *When the F market maker passively transacts at the F offer at time t, she will seek to aggressively hedge herself by transacting some quantity at the S offer at time $t+\tau$ if the expected S offer price at $t+\tau$ is below the F mid price at t: $\mathbb{E}\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right] < p^F_{m,t}$.*

2. *When the F market maker passively transacts at the F bid at time t, she will seek to aggressively hedge herself by transacting some quantity at the S bid at time $t+\tau$ if the expected S bid price at $t+\tau$ is above the F mid price at t: $\mathbb{E}\left[p^S_{m,t+\tau} - \xi^S_{t+\tau}\right] > p^F_{m,t}$.*

3. *When the S market maker passively transacts at the S offer at time t, she will seek to aggressively hedge herself by transacting some quantity at the F offer at time $t+\tau$ if the expected F offer price at $t+\tau$ is below the S mid price at t: $\mathbb{E}\left[p^F_{m,t+\tau} + \xi^F_{t+\tau}\right] < p^S_{m,t+1}$.*

4. *When the S market maker passively transacts at the S bid at time t, she will seek to aggressively hedge herself by transacting some quantity at the F bid at time $t+\tau$ if the expected F bid price at $t+\tau$ is above the S mid price at t:* $\mathbb{E}\left[p^F_{m,t+\tau} - \xi^F_{t+\tau}\right] > p^S_{m,t+1}$.

Proof. When F and S market volatilities are equal, Equation (59) and its analog for the S market maker, reduce to

$$q^S_{m,t+\tau} = \left(p^F_{m,t} - \mathbb{E}_m\left[p^S_{m,t+\tau} + \xi^S_{t+\tau}\right]\right)\left(\frac{\partial p^S_{m,t+\tau}}{\partial q^S_{m,t+\tau}}\right)^{-1} \tag{67}$$

$$q^F_{m,t+\tau} = \left(p^S_{m,t} - \mathbb{E}_m\left[p^F_{m,t+\tau} + \xi^F_{t+\tau}\right]\right)\left(\frac{\partial p^F_{m,t+\tau}}{\partial q^F_{m,t+\tau}}\right)^{-1}. \tag{68}$$

By inspection, we observe that $q^S_{m,t+\tau}$ and $q^F_{m,t+\tau}$ are only positive when the conditions of the proposition are met. □

In the foregoing development of the model, we have neglected the treatment of rebates and transactions costs. Since fee structures vary widely across exchanges and markets, it is difficult to make uniform statements as to how they would affect our stated results. In the markets used for our analysis in Section 13.3, both market makers and aggressive takers pays fees to transact futures contracts at the CME, while market makers receive rebates and aggressive takers pay fees to trade equities at the Nasdaq. Let ϕ^F denote the fee to transact in the futures market and ϕ^S and ρ^S the taker fee and maker rebates in the spot market, respectively. For the investor, $-2\phi^F$ and $-2\phi^S$ are added to numerators of Equations (53) and (54), respectively. For the F maker, who pays fees to transact in both markets, $-\phi^F - \phi^S$ is added to the numerator of Equation (59), which serves to tighten the arbitrage condition: the F market maker will demand an even lower offer price in the S market in order to exploit a potential cross-market arbitrage. On the other hand, for the S market maker, $\rho^S - \phi^F$ is added to the numerator of Equation (59), since the S maker earns a rebate for passively providing liquidity in the S market and pays a fee to exploit an arbitrage profit in the F market. Depending on the net effect ($\rho^S > \phi^F$ or $\rho^S < \phi^F$) the fees may serve to relax or tighten the required difference in prices across market to make a profitable arbitrage. For the CME and Nasdaq in particular, this difference is likely to be positive, where the fee to trade a single ES futures contract is \$0.25 (for the most active market makers, who have access to lowest latency communications technology) and the rebate to trade 500 SPY shares (the equivalent of one ES contract – see Section 13) is roughly \$1.05. The net difference, \$0.80, however, is almost five times smaller than the difference between the two half spreads, \$3.75, and so we do not deal with it here.

12 Econometric Methodology

To test the implications of our model, we need an econometric measure of information flow and price responsiveness across markets. Hasbrouck (1995) introduced an econometric methodology which has become the backbone for measuring information shares of distinct markets trading a single asset. Formally, given N markets and an $N \times 1$ vector of prices in those markets, $\boldsymbol{p}_t$, Hasbrouck (1995) assumes that the prices are individually nonstationary and that the vector is cointegrated of order $N-1$. Expressing the cointegrated system with the common trends representation of Stock and Watson (1988),

$$\boldsymbol{p}_t = \boldsymbol{p}_0 + \boldsymbol{\psi} \left(\sum_{s=1}^{t} \boldsymbol{e}_s \right) \boldsymbol{\iota} + \Psi^*(L)\boldsymbol{e}_t, \tag{69}$$

Hasbrouck (1995) defines the information share of market j to be

$$S_j = \frac{\psi_j^2 \Omega_{jj}}{\boldsymbol{\psi}\Omega\boldsymbol{\psi}'}, \tag{70}$$

where $\boldsymbol{e}_t$ is a vector of price innovations to each of the markets, $\sum_{s=1}^{t} \boldsymbol{e}_s$ is the common random walk component to prices, $\boldsymbol{\psi}$ is a row vector which measures the long-run impact of innovations $\boldsymbol{e}_t$ on prices (which is taken from the moving average representation of price differences), and Ω is the covariance matrix of innovations. In essence, this measure of information share is variance decomposition: Equation (70) equates the share of information in market j to the total variance contribution of its price innovation. While Equation (70) assumes a diagonal covariance matrix, Ω, Hasbrouck (1995) also computes bounds on the variance decomposition by using a Cholesky factorization for non-diagonal covariance matrix.

The aforementioned measure of information share has been widely adopted because of its elegance of interpretation and its parsimonious representation as a cointegrated time series model. However, it suffers from several weaknesses. First, to maintain synchroneity across elements of the price vector, prices must be observed at identical times. In practice, this means that they must be observed in clock time. Brada et al. (1966), Mandelbrot and Taylor (1967), Clark (1973), and more recently Ane and Geman (2000) and Aldrich et al. (2015) all demonstrate the advantages of expressing prices at granular time intervals using some measure of (subordinated) transaction time. In particular, Aldrich et al. (2015) highlights the importance of expressing price evolution through a dynamic model of transaction arrival. Second, it requires a parametric, linear time series model and a set of assumed cointegrating relationships. Third, as high-frequency time series data becomes increasingly rich, the time series model becomes increasingly large and complex. To deal with overparamterization in the model, it is necessary to impose parameter restrictions, such as the polynomial smoothing of coefficients performed by Hasbrouck (1995). Finally, the methodology only utilizes price information and neglects a second piece of information that is attributed to each trade and quote: size. Incorporating volume and order flow (defined as the difference between volume done on the best offer and volume done on the best bid) is essential to understanding total market share of information.

We propose a measure of information flow that is non-parametric and designed explicitly for high-frequency, transaction-time data. It is also well suited to measuring the model implications of Proposition 11.8. This econometric measure correlates events (e.g., transactions or changes in displayed liquidity) in two markets and measures deviations from a null hypothesis of the events in one market, under specific conditions, having no informative impact on concurrent or subsequent activity in another market. The measure inherently accounts for volume and order flow and also flexibly accounts for communication latency between markets.

Formally, let $\mathcal{X}$ and $\mathcal{Y}$ be the sets of all possible events that can occur in a conditioning market, X, and a responding market, Y, at any time t. Suppose we are interested in understanding how the events in a subset $\mathcal{X}^* \subseteq \mathcal{X}$ affect events in a subset $\mathcal{Y}^* \subseteq \mathcal{Y}$. To do so, we define,

$$\delta_s = \mathbb{E}\left(y_{t+s}|x_t\right) - \mathbb{E}\left(y_{t+s}\right), \quad \text{for} \quad s = 0, 1, \ldots, S, \tag{71}$$

for $x_t \in \mathcal{X}^*$ and $y_{t+s} \in \mathcal{Y}^*$ for $s = 0, 1, \ldots, S$. That is, $\{\delta_s\}_{s=0}^S$ simply measures the conditional effect of events in $\mathcal{X}^*$ on events in $\mathcal{Y}^*$ over horizon S. We estimate $\{\delta_s\}_{s=0}^S$ via simple frequency counts:

$$\hat{\delta}_s = \frac{1}{N_{x^*}} \sum_{i=1}^{N_{x^*}} \mathbb{1}(y_{t_i+s}|x_{t_i}) - \mathbb{E}\left(y_{t+s}\right), \tag{72}$$

where N_{x^*} is the number of events in $\mathcal{X}^*$ that occurs during the period of interest and where we have assumed that the unconditional expectation of y_{t+s} is known *a priori*. In the case that the unconditional expectation of y_{t+s} is unknown, it may also be estimated through a similar frequency count, by sampling an identical number, N_{x^*}, of events during the period of interest. The estimator of Equation (72) amounts to computing the difference of two histograms.

In Section 13, we will utilize the estimator $\{\hat{\delta}_s\}_{s=0}^S$ in the following manner. We will separately consider two cases for the originating market, $X \in \{ES, SPY\}$, with the responding market, Y, being the other. Given a choice of X, we will focus on distinct subsets of market events, $\mathcal{X}^*$: all possible transactions (sizes) on one side of the order book (bid or offer), where the transacted price at time t is either above or below the displayed quotation on the same side of the order book in the responding market at time $t + s$, for $s \leq S$. For example, with $X = ES$, one of the four subsets would be transactions on the ES bid, when that bid price is below the quoted bid in the SPY market 7 ms later. Note that as we have defined it, the number of events in the conditioning set, N_{x^*}, is dependent on lag s. In this case, $\{x_{t_i}\}_{i=1}^{N_{x^*}}$ represents the *observed* bid transactions in the ES during a single day that abide by the aforementioned restriction. The result is eight different collections of estimators: $\left\{\delta_s^{X,p,d}\right\}_{s=0}^{S}$ where $X \in \{ES, SPY\}$, $p \in \{b, o\}$ (bid or offer) and $d \in \{\downarrow, \uparrow\}$ (below or above). $\mathcal{Y}^*$ will represent the set of all possible values of order flow, or differences (in size) of transactions done on the offer and bid, in the responding market per unit of volume in the originating market, and $\left\{\{y_{t_i+s}\}_{i=1}^{N_{x^*}}\right\}_{s=0}^{S}$ will represent the *observed* differences in sizes (per unit $\{x_{t_i}\}_{i=1}^{N_{x^*}}$) following each observed event in the originating market. We make the assumption that $\mathbb{E}[y_t] = 0 \ \forall t$, which states that unconditional expectation of transacted size on the offer is equal to that on the bid. A simple empirical check has shown this to be a very reasonable assumption.

We maintain the following null hypothesis:

Hypothesis 12.1. *Events in the originating market do not informatively affect the distribution of subsequent events in the responding market:* $\delta_s^{X,p,d} = 0$ *for all possible choices of* (X, p, d).

As discussed in Section 10, conventional wisdom suggests that events in the ES market subsequently inform events in the SPY market. Thus, a naive assumption would be that $\delta_s^{ES} \neq 0$ for some $s \in \{0, \ldots, S\}$, but that $\delta_s^{SPY} = 0 \ \forall s$ (note that we have dropped the superscripts b and d since the stated assumption would hold for all values). Proposition 11.8, however, provides a set of testable predictions, which are sometimes incongruent with the stated conventional wisdom. We consolidate these testable predictions in the following proposition.

Proposition 12.2.

1. $-1 \times \delta_s^{ES,b,\downarrow} > 0$ *for some* s.
2. $\delta_s^{ES,b,\uparrow} = 0$ *for all* s.
3. $\delta_s^{ES,o,\uparrow} > 0$ *for some* s.
4. $\delta_s^{ES,o,\downarrow} = 0$ *for all* s.
5. $-1 \times \delta_s^{SPY,b,\downarrow} > 0$ *for some* s.
6. $\delta_s^{SPY,b,\uparrow} = 0$ *for all* s.
7. $\delta_s^{SPY,o,\uparrow} > 0$ *for some* s.

8. $\delta_s^{SPY,o,\downarrow} = 0$ *for all* s.

Note that in two cases we have reversed the sign of the estimator in order to consistently focus on positive deviations from the null hypothesis. This will assist visualization and aggregation of our empirical results in Section 13.

Figure 24 displays a stylized representation of δ_s in statements 1,3,5 and 7 of Proposition 12.2. If the figure conformed exactly to the model in Section 11, δ_s would be a dirac function at $s = \tau$, the inter-market communication latency, with height $q_m^{X^c}$, the number of shares that the market maker chooses to route to the responding market. Instead, we have depicted the response function as we might expect it to appear in a market with a heterogeneous group of market makers who have access to different communication technology and different ability to assess and compute the relative information across markets. In the next section, we will empirically estimate δ_s for each of the subcases considered in Proposition 12.2.

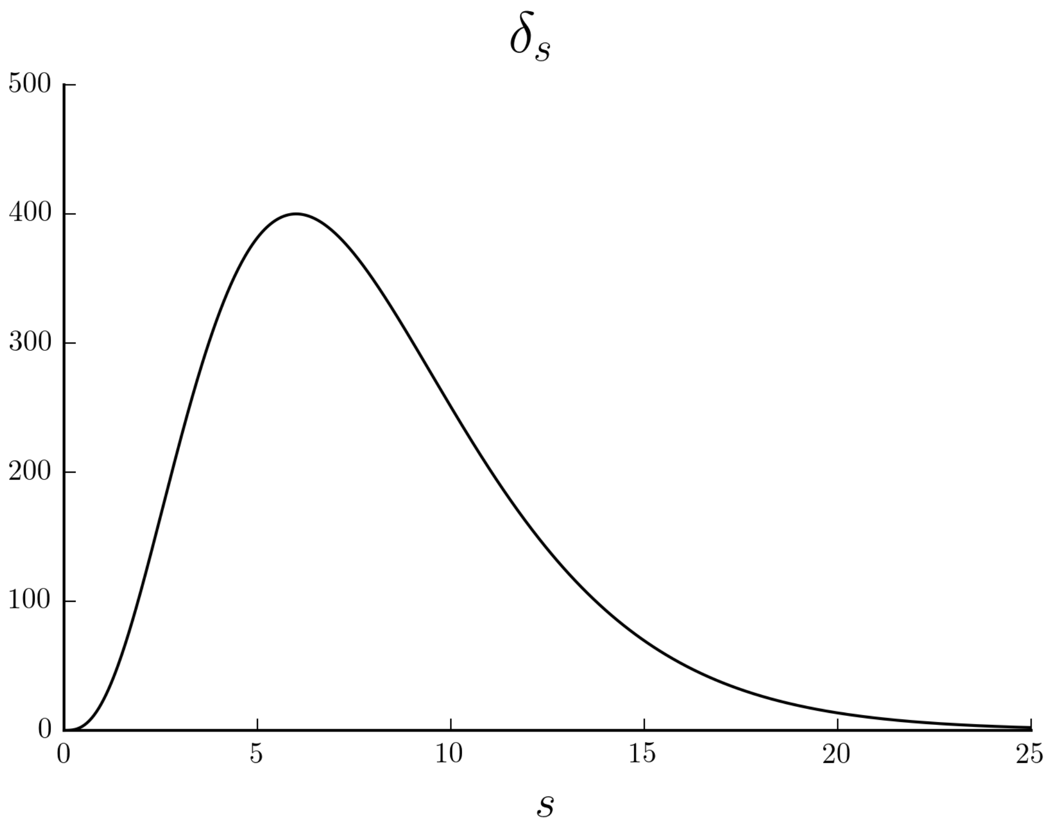

Figure 24: Stylized representation of information response measure.

13 Empirical Application

We now apply the econometric methodology of the last section to empirically test the model implications of Section 11. We begin by describing our data and the necessary adjustments we make in order to equate the prices of futures and equity assets.

13.1 Data

Our data comprise all transactions for the E-mini S&P 500 futures contract (CME Group symbol ES, commonly known as the E-mini) and the State Street Global Advisers SPDR S&P 500 exchange traded fund (NYSE Arca symbol SPY) between Jun 16, 2014 and Sep 11, 2014. The E-mini trades exclusively at the CME Group Globex matching engine in Aurora, Illinois, longitude -88.24° W, latitude 41.80° N. The market consists of five listed contracts at all times, expiring on the March quarterly cycle, with expiry occurring on the third Friday of the designated month. The bulk of trading interest resides in the near-month contract, although common practice dictates that interest shifts to the second-month contract exactly one week prior to expiry. This transition, on the second Friday of the expiry month, is known as the *roll date*. The sample period for our data was selected to coincide with all trading days between the two roll dates for the September 2014 ES contract (symbol ESU4). The E-mini market is open nearly continuously each week from Sunday, 6:00 p.m. ET to Friday, 5:00 p.m. ET, aside from a daily trading halt from 4:15 – 4:30 p.m. ET and a daily maintenance period from 5:00 – 6:00 p.m. ET. Although trading occurs during all market hours, the majority of activity coincides with U.S. equities market hours, 9:30 a.m. – 4:00 p.m. ET, Monday through Friday. For this reason, and because we are interested in correlating with activity in the equities market, we restrict attention to U.S. equities market hours. The contract is quoted in S&P 500 index points, with a minimum spread of 0.25 index points, although the notional value is \$50× the index. Panel (a) of Table 2 reports basic price and transaction size summary statistics for the period we consider; for example, the average price of the contract was 1959.40 index points, corresponding to a notional value of \$97,970.00.

(a) Price and Size Summary Statistics							
Asset	# Trades	P_{min}	P_{max}	$\bar{P}$	S_{min}	S_{max}	$\bar{S}$
ES	16,448,563	1899.75	2010.00	1959.40	1	1792	3.826
SPY	2,781,812	190.55	201.58	196.43	1	92,600	309.57
(b) Bid/Offer Summary Statistics							
Asset	# Bid Trades	P^b_{below}	P^b_{above}	# Offer Trades	P^a_{below}	P^a_{above}	
ES	1,829,316	1,159,262	670,054	1,815,467	692,672	1,122,795	
SPY	615,919	137,516	478,403	517,517	411,338	106,179	
(c) Integrated Order Flow Response (units = SPY shares)							
Market	$s = 0$	$s = 5$	$s = 10$	$s = 15$	$s = 20$	$s = 25$	$s = 30$
ES	0.04564	4.811	9.532	10.83	11.88	12.72	13.21
SPY	1.888	89.01	640.1	1135	1260	1335	1397
(d) Integrated Order Flow Response Value (dollars)							
Market	$s = 0$	$s = 5$	$s = 10$	$s = 15$	$s = 20$	$s = 25$	$s = 30$
ES	633.31	66,760.18	132,260.25	150,301.39	164,804.12	176,503.71	183,258.99
SPY	230.71	10,877.46	78,223.08	138,747.88	153,994.15	163,184.48	170,688.22

Table 2: Panel (a): Summary statistics for ES and SPY transaction prices (P) and sizes (S) between Jun 16 and Sep 11, 2014. Panel (b): Summary statistics for ES and SPY bid/offer transaction prices relative to bid/offer quotes in the other market. Panel (c): Integrated order flow responses for bids and offers in each originating market. Panel (d): Valuation of integrated order flow responses for bids and offers in each originating market.

The SPY exchange traded fund (ETF) is listed with NYSE Arca and trades on U.S. equities exchanges. We confine attention to direct feed data obtained from the Nasdaq matching engine in Carteret, New Jersey (longitude -74.25° W, latitude 40.58° N). While it would be desirable to obtain a transaction record for SPY activity across all exchanges (such as from the consolidated tape), we emphasize the importance of using direct feed data, which has more accurate time stamps and which does not include an additional reporting delay to the Securities Information Processor (SIP). Consolidated tape data (e.g. NYSE Daily TAQ) includes the SIP delay and would introduce ambiguity in our cross-market latency analysis. Table 3 lists the average market share (and standard deviation) of SPY transactions conducted on each equities exchange for all trading days in Apr, 2014, the most recent month for which we have consolidated tape data. The table shows that Nasdaq represents nearly 25% of all SPY market volume and is the largest exchange in terms of SPY market share. As noted above, equities markets are open from 9:30 a.m. – 4:00 p.m., Monday through Friday. Although some before- and after-market trading occurs for liquid equities, the vast majority of trading occurs during regular market hours, and for this reason we confine attention to those hours. The SPY ETF is priced at 1/10th the value of the S&P 500 index, and per Rule 612 of Reg NMS ("the rule"), market makers are prohibited from displaying or accepting quotations priced in an increment smaller than \$0.01. Panel (a) of Table 2 reports basic price and transaction size summary statistics for SPY during our sample period.

Exchange	Mean	Std. Dev
Nasdaq	0.234	0.031
NYSE Arca	0.210736	0.018592
FINRA	0.163565	0.015833
BATS	0.157954	0.013971
DirectEdge X	0.061514	0.005623
BATS Y	0.056861	0.006155
Nasdaq BX	0.054256	0.025917
DirectEdge A	0.043766	0.004578
Nasdaq PSX	0.013788	0.001938
NSX	0.001936	0.000586
CSE	0.001307	0.000401
CBOE	0.000022	0.00001

Table 3: Market share (by number of transactions) for all participating equities exchanges during Apr, 2014.

Given that the E-mini and SPY ETF trade the same underlying quantity, it is expected that their markets are tightly linked. Laughlin et al. (2014) highlights the correlation of shifting liquidity from E-mini futures to equities markets, using an empirical measure similar to what we propose in Section 12. Their work further documents that microwave communication infrastructure has reduced the latency of information transmission to just over 4 milliseconds (ms), only slightly more than the 3.93 ms theoretical minimum to transmit messages in a frictionless environment on the great circle between Aurora and Carteret. An important distinction to make between the two assets is their relative notional values: since the E-mini is valued at \$50× the index and the SPY is valued at 1/10th the index, one contract of the former is worth roughly 500 shares of the latter. Further, after accounting for the order of magnitude difference in price quotations between the two assets, the minimum increment of 0.25 index points for the E-mini is 2.5 times larger than the \$0.01, or 0.10 index point, minimum increment for the SPY. This latter relationship is of crucial importance to our empirical application, as it represents the channel of information transmission that we highlighted in Section 11.

13.2 Basis Adjustment

In order to empirically test the statements of Propositions 11.8 and 12.2, we need to adjust the prices of the ES and SPY contracts so that they are comparable. The standard futures pricing equation is

$$F_t = e^{(r_f - \rho)(T-t)} S_t, \tag{73}$$

where F_t is the futures price, S_t is the spot price of the underlying asset (in this case the S&P 500 index), r_f is the risk-free interest rate, ρ is the dividend rate and $T - t$ is the time until expiry of the futures contract. Equation (73) shows that differences in futures and spot prices arise from stochastic variations in r_f and ρ, as well as the deterministic movement of time, $T - t$. The conventional way to quantify the wedge between prices is via the basis,

$$b_t = F_t - S_t = \left(e^{(r_f - \rho)(T-t)} - 1 \right) S_t, \tag{74}$$

which can be accurately estimated since the stochastic fluctuations in r_f and ρ are typically quite small over short time horizons. Rather than working in a forward fashion, estimating interest and dividend rates at a daily frequency, we introduce a methodology to infer what the basis must be from the empirical distribution functions of ES/SPY bid and offer differences. Figure 25 displays histograms of ES/SPY best bid differences (blue) and best offer differences (red) at the times of all transactions in the SPY market on Aug 4, 2014. The histogram bars are centered on the unique values of the differences (i.e. the bars do not aggregate values on the x-axis) and demonstrate that there are very few unique differences on Aug 4, 2014. This latter fact is owing to the price discreteness of both assets.

To infer the basis for a specific day, we must first understand the nature of *adjusted* prices. We begin with the following assumption.

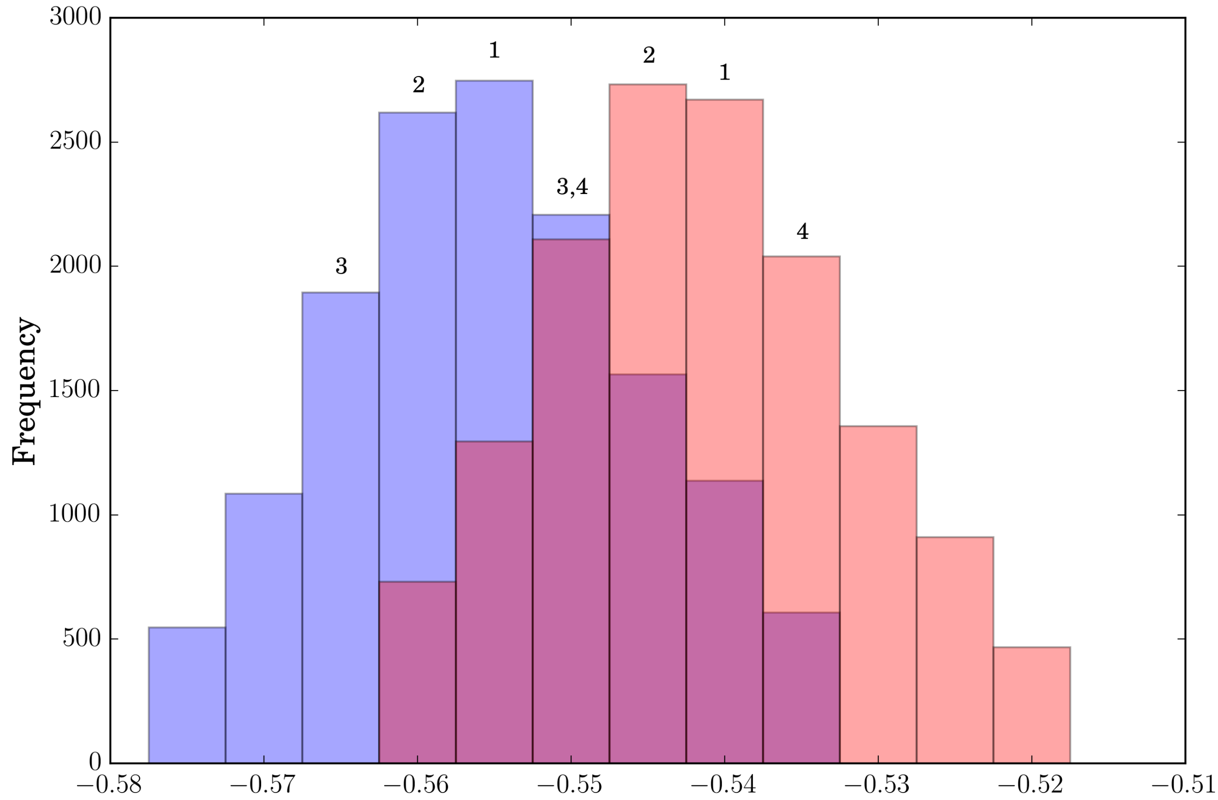

Figure 25: Histograms of ES/SPY bid differences (blue) and ES/SPY offer differences (red) for Aug 4, 2014.

Assumption 13.1. *On average, basis-adjusted bids and offers for the ES and SPY are symmetrically quoted around a latent, fair price for the S&P 500 index.*

Combining Assumption 13.1 with the fact that the ES price increment is 2.5 times larger than that of the SPY, we conclude that the SPY bid and offer will typically sit completely inside of the ES bid and offer. Panel (b) of Table 2 separates all ES and SPY transactions according to whether they were done on the bid or offer, and reports the total counts for which basis-adjusted (described below) bid/offer transactions in each market are above or below the best bid/offer quote in the other market. The data corroborate our conclusion: at the time of an ES transaction, roughly two-thirds of ES prices sit outside the best quotes in the SPY market and at the time of a SPY transaction, roughly 80% of SPY prices sit inside the best quotes in the ES market. Although the values reported in Table 2 do not account for inter-market communication latency, we have computed them for time shifts of $\tau \leq 5$ ms and verified that the aggregate numbers change very little. We also note that the number of ES and SPY transactions in panel (b) do not equal the totals in panel (a); this is due to the fact that many transactions occur in the same millisecond (a result of order splitting) and because panel (b) only considers unique transaction times.

Figure 26 depicts several scenarios when the basis-adjusted ES bid and offer are quoted at 200.00 and 200.025. While our choice of price grid in the figure only allows the ES/SPY bids to perfectly align, a symmetric example exists in which the offers perfectly align (e.g. when the ES is quoted at 200.025 and 200.05). Assumption 13.1 states that Case 3 (and its symmetric analog) in the diagram is most common: with highest probability, the SPY bid is either \$0.005 or \$0.01 greater than the ES bid, and, respectively, the SPY offer is \$0.01 or \$0.005 less than the ES offer. These two scenarios correspond to the histogram bars labeled '1' and '2' in Figure 25. The next most probable cases (Case 2 in Figure 26 and its analog) occur when bids or offers align, which correspond to the histogram bars labeled '3' and '4'. The central overlapping bars (labeled '3,4') identify the basis: this is the price shift which causes the bids and offers to exactly align. For Aug 4, 2014, we conclude that the basis was -0.55. Repeating this procedure for each day in our dataset, we construct an inferred basis series, which is used to align prices for our econometric test. We view the resulting aligned prices as reasonable inputs for cross-market comparison since the intra-day basis market is very active for traders that participate in the ES and SPY markets.

In practice, a single basis does not exist since hedging across markets and products entails a net lending or borrowing position; a small interest rate spread applies to borrowers (traders that are short ES and long SPY) and lenders (traders that are long ES and short SPY). Although the spread is trader specific and dictated by the market for the basis, we take a commonly reported series of basis spreads (ranging from 0.065 on Jun 16, 2014 to 0.001 on Sep 11, 2014) from Bloomberg and further adjust ES prices upward to account for positions that are net borrowing (a passive fill on the ES offer and aggressive hedge on SPY offer or a passive fill on

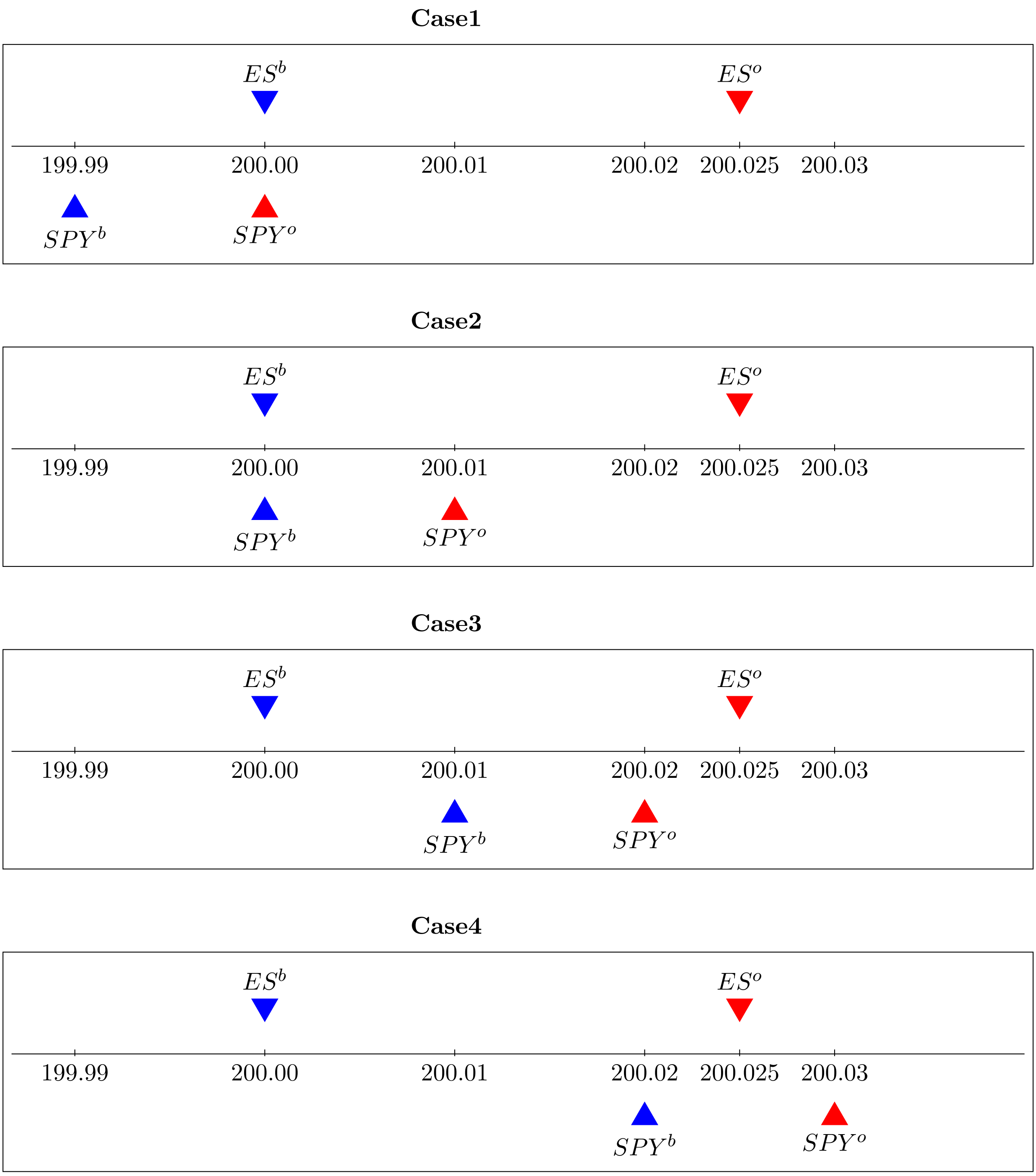

Figure 26: Examples of *adjusted* relative ES and SPY prices.

SPY bid and aggressive hedge on ES bid) and downward to account for net lending (a passive fill on the SPY offer and aggressive hedge on ES offer or a passive fill on ES bid and aggressive hedge on SPY bid).

13.3 Results

To test the implications of our model, we use basis-adjusted prices to determine the relationship of ES/SPY bids and offers at the times of all transactions in each market. Since the data are not contiguous across days, we separately estimate the values of $\delta^{X,p,d} = \left\{\delta_s^{X,p,d}\right\}_{s=0}^{S}$ in Proposition 12.2 for each day in our dataset. Recall that $\delta_s^{X,p,d}$ represents the difference in number of shares transacted at the offer and bid in the responding market, per unit of volume transacted in the originating market (on bid or offer, depending on $p \in \{b, o\}$) when the prices in the two markets abide by restriction d at time lag s. We emphasize the importance of comparing prices at distinct time lags, s: it is essential for estimation purposes to account for hypothetical communication latency when comparing bids and offers across markets. For example, conditional on a transaction at the ES bid at time t, $\delta_s^{ES,b,\downarrow}$ must be estimated only for SPY bid transactions at time $t+s$ when $P_t^{ES,b} < P_{t+s}^{SPY,b}$ and not necessarily when $P_t^{ES,b} < P_t^{SPY,b}$. In doing this, we do not rigidly assume a value for communication latency, τ, but instead allow the communication latency to be observed in the econometric estimates. As noted in Section 13.1, we expect $\tau \approx 4$ ms, which would imply $\delta_s^{X,p,d} \approx 0$ for $s < 4$ ms in all cases outlined in Proposition 12.2.

We make two more refinements on the sets of conditioning events. In order to ensure that the estimated order flow responses, $\hat{\delta}^{X,p,d}$, are not partially attributed to information prior to their conditioning events, we only consider transactions in each market that are not preceded by transactions in the 7 ms immediately prior. This threshold is enough to ensure that all information from both markets has been impounded into the conditioning event price, and that the estimated responses are not an artifact of information that was available before the event. In addition, for the sets of SPY conditioning transactions, we only consider transactions of 500 shares or more. Since a single ES contract is equivalent to 500 SPY shares, a SPY market maker that is filled on bid or offer for less than 500 shares would have to reverse her net position in the ES market by taking advantage of a potential arbitrage opportunity. These two refinements reduce the total sample size from 1,829,316 ES events and 615,919 SPY events during our sample period (reported in panel (b) of

Table 2) to 1,145,481 and 112,874 events, respectively.

Figures 27 and 28 show daily estimates of the eight order flow response estimators listed in Proposition 12.2. Specifically panels (a), (b), (c) and (d) of Figure 27 show bid responses, $\delta^{ES,b,\downarrow}$, $\delta^{ES,b,\uparrow}$, $\delta^{SPY,b,\downarrow}$ and $\delta^{SPY,b,\uparrow}$, respectively, and panels (a), (b), (c) and (d) of Figure 28 show offer responses, $\delta^{ES,o,\uparrow}$, $\delta^{ES,o,\downarrow}$, $\delta^{SPY,o,\uparrow}$ and $\delta^{SPY,o,\downarrow}$, respectively. The bold, solid lines in each panel represent the median of daily response values and dotted lines correspond to 0.05 and 0.95 empirical quantiles of the same. All y-axis units have been scaled to units of SPY shares (i.e. ES responses have been scaled by 500). As described in Section 13, the order flow response estimators, $\delta^{X,p,d}$, are computed point by point, for each millisecond following a bid or offer transaction in an originating market. Specifically, the number of conditioning events N_{x^*} is determined by the number of originating events, N_x, that satisfy specific relationships of relative bids and offers across markets, and changes for each millisecond time interval following an event in the originating market. In Figures 27 and 28 we have rescaled the all values by $\frac{N_{x^*}}{N_x}$, where N_x is the common number of events in the originating market; e.g. ES bid events. Thus, prior to the rescaling, the order flow response for an ES bid transaction, when that bid is below a subsequent SPY bid, is interpreted as $\delta_s^{ES,b,\downarrow} = \mathbb{E}\left[SPY \text{order flow} | ES_t^b < SPY_{t+s}^b\right]$, whereas after the rescaling it is interpreted as $\delta_s^{ES,b,\downarrow} = \mathbb{E}\left[SPY \text{order flow} | ES_t^b < SPY_{t+s}^b\right] Pr\left(ES_t^b < SPY_{t+s}^b\right)$. Given a symmetric interpretation of $\delta_s^{ES,b,\uparrow}$, the sum of the rescaled responses is interpreted as $\delta_s^{ES,b,\downarrow} + \delta_s^{ES,b,\uparrow} = \mathbb{E}\left[SPY \text{order flow} | ES^b\right]$. The result is that the summation of response estimates in panels (a) and (b) or panels (c) and (d) of Figures 27 and 28 are estimates of order flow following bid or offer events in the origination markets, regardless of relative price conditions.

Panels (a) in both figures depict the expected, conventional response: when a trade occurs on the ES bid or offer, there is a subsequent preponderance of trading on the SPY bid and offer, respectively. Further, these panels clearly identify the cross-market communication latency to be approximately 4 or 5 ms, as the order flow response estimates do not appear to be statistically different from zero for

$s \leq 3$ ms. Panels (b) in both figures, however, demonstrate that when a transaction occurs at the ES bid or offer under conditions that our model suggests are not profitable in terms of cross-market arbitrage ($P_t^{ES,b} \geq P_{t+s}^{SPY,b}$ or $P_t^{ES,o} \leq P_{t+s}^{SPY,o}$) the statistical significance of the response disappears. This latter result directly contradicts conventional wisdom that information in the futures market precedes that of the spot market.

The lower rows of Figures 27 and 28 show the responses of the ES market to transactions on the bid and offer in SPY. Similar to panels (a), panels (c) depict a strong order flow response in the ES market when a trade occurs in the SPY market under favorable cross-market arbitrage conditions: $P^{SPY,b} < P^{ES,b}$ and $P^{SPY,o} > P^{SPY,o}$. Panels (d), however, show that when these conditions do not hold, the significance of the response disappears, as predicted by the model. Once again, while panels (d) are congruent with unconditional behavior across the ES/SPY market, panels (c) violate convention and strongly support our model.

An interesting feature of panels (c) is that the peak of the response occurs much later than that of panels (a) – closer to 10 ms than 4 ms. This is somewhat unexpected, as the microwave lines allow for full duplex transmission rates, both for the purpose of providing reverse information flow, as well as for the basic need to deliver fast message confirmations in the traditional, Aurora to New Jersey, direction. The delayed peak in the New Jersey to Aurora direction could arise because the bulk of traders exploiting the reverse cross-market arbitrage are less sophisticated, and are somewhat slower, or because they are more likely to use fiber optic infrastructure instead of microwave infrastructure. Despite this somewhat delayed response, it is clear that events in the SPY market have a significant impact on trading in the ES market when cross-market arbitrage conditions create profit opportunities for market makers.

The final important feature to note from Figures 27 and 28 is that the responses in panels (c) are almost two orders of magnitude larger than those of panel (a). This is partly related to the relative coarseness of ES/SPY contract sizes. When an ES market maker is passively filled on one or multiple contracts, the relative

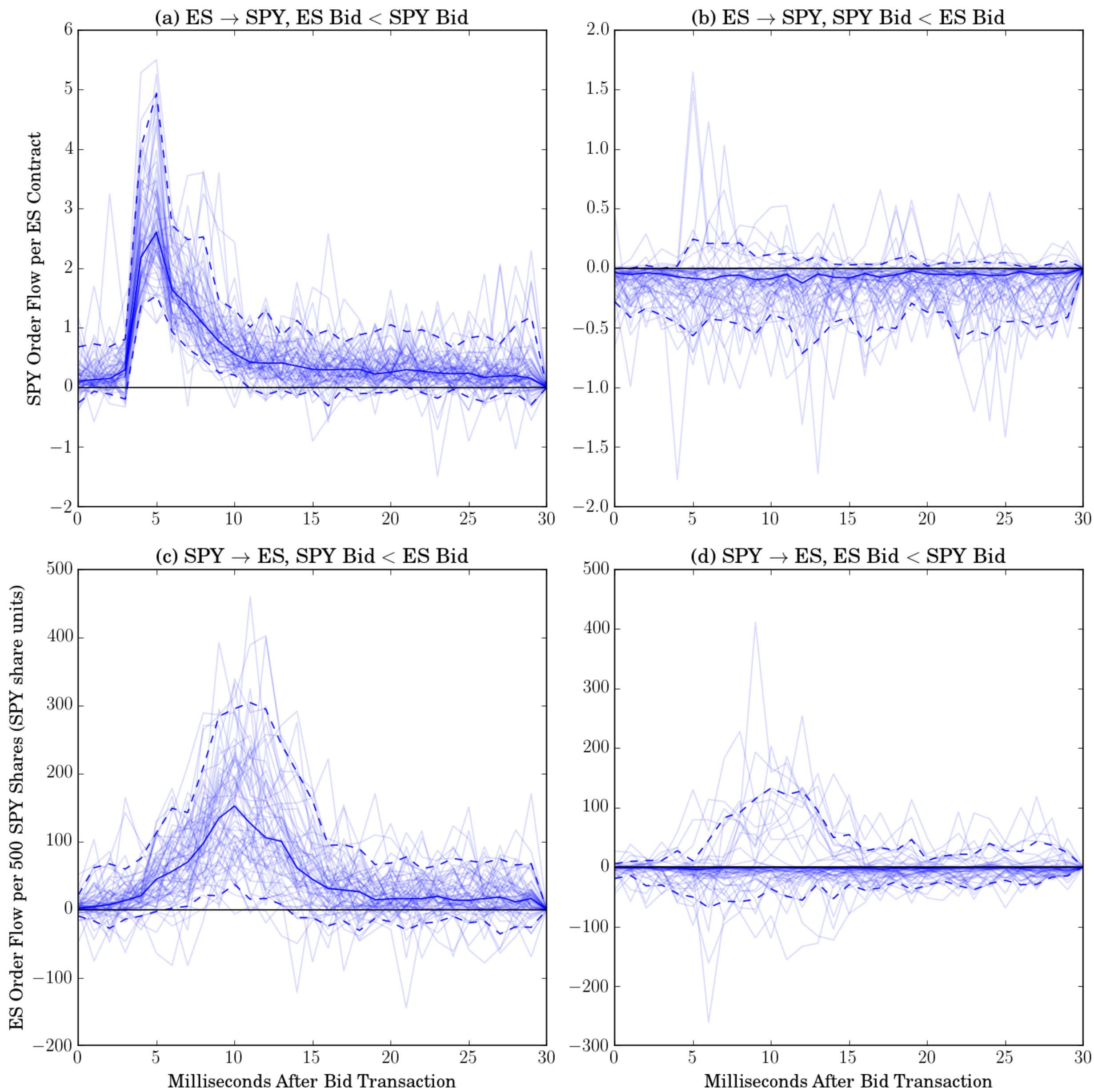

Figure 27: Order flow responses for ES and SPY bid transactions. Panel (a) shows SPY order flow response after ES bid event, when $P_t^{ES,b} \leq P_{t+s}^{SPY,b}$. Panel (b) shows the same response but for $P_t^{ES,b} \geq P_{t+s}^{SPY,b}$. Panels (c) and (d) display analogous responses of the ES (expressed in units of SPY shares) after SPY bid transactions for the (respective) cases of $P_t^{ES,b} \geq P_{t+s}^{SPY,b}$ and $P_t^{ES,b} \leq P_{t+s}^{SPY,b}$.

granularity of SPY share sizes gives her a high degree of flexibility in choosing her hedge quantity. In the reverse scenario, however, SPY market makers are forced to discretely round up or down to multiples of 500 SPY share equivalents. As a whole, the effect of relative contract size contributes to an overall larger impact of

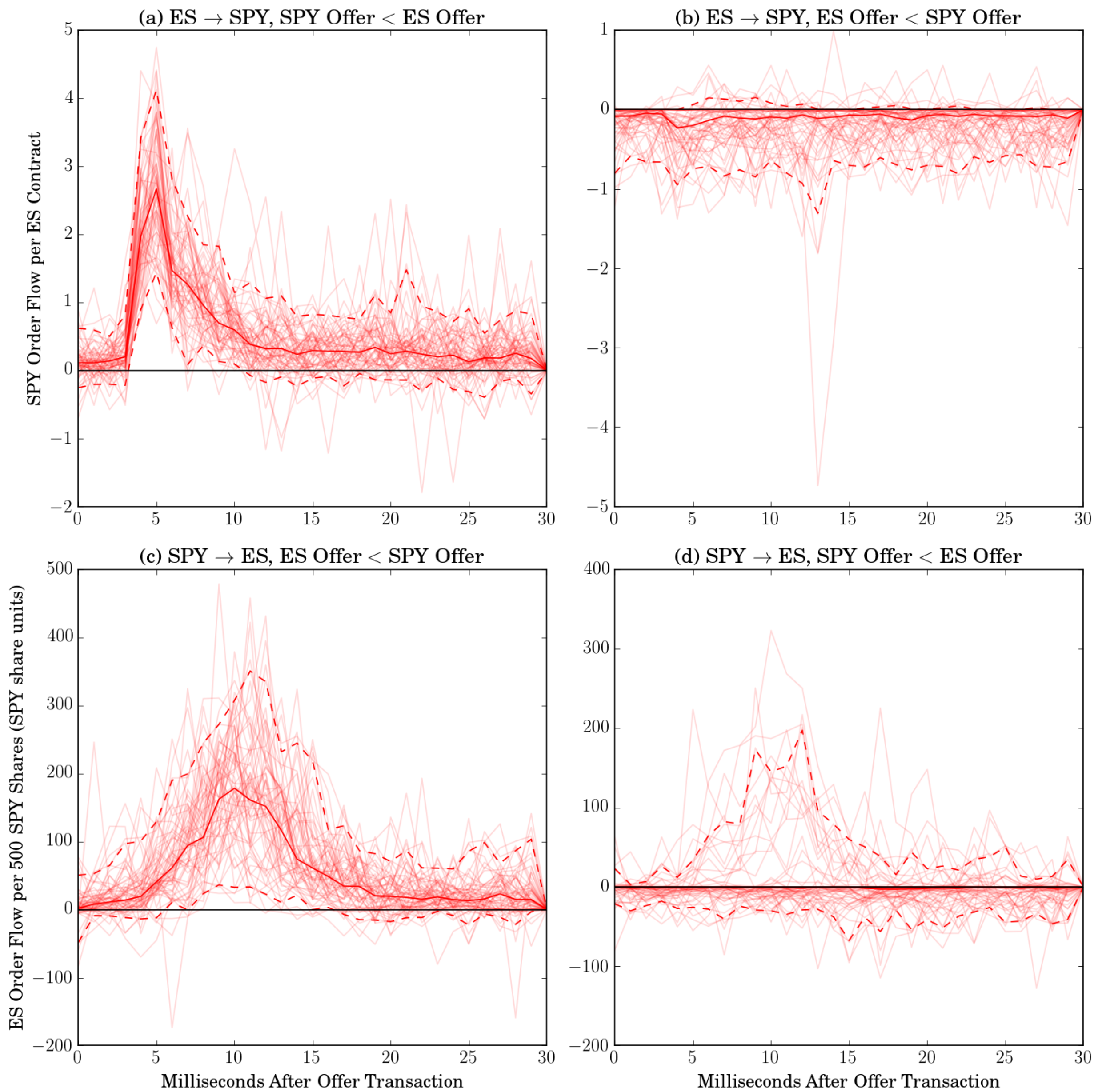

Figure 28: Order flow responses for ES and SPY offer transactions. Panel (a) shows SPY order flow response after an ES offer event, when $P_t^{ES,o} \geq P_{t+s}^{SPY,o}$. Panel (b) shows the same response but for $P_t^{ES,o} \leq P_{t+s}^{SPY,o}$. Panels (c) and (d) display analogous responses of the ES (expressed in units of SPY shares) after SPY offer transactions for the (respective) cases of $P_t^{ES,o} \leq P_{t+s}^{SPY,o}$ and $P_t^{ES,o} \geq P_{t+s}^{SPY,o}$.

SPY events on subsequent ES trading. Thus, although the conditions under which SPY events informatively lead ES transactions are relatively infrequent, they are individually more informative.

We summarize the estimated order flow responses in panel (c) of Table 2 by integrating the curves generated by the medians depicted in Figures 27 and 28. Specifically, the integrated order flow response for originating market $X \in \{ES, SPY\}$is defined as

$$\Delta_s^X = \int_0^s 0.5\left(\delta_s^{X,b,\downarrow} + \delta_s^{X,b,\uparrow}\right) + 0.5\left(\delta_s^{X,o,\downarrow} + \delta_s^{X,o,\uparrow}\right) ds \tag{75}$$

Given our prior interpretation of $\delta_s^{X,b,\downarrow} + \delta_s^{X,b,\uparrow}$, Δ_s^X represents the total order flow response in the responding market over horizon s, for any transaction (bid or offer) in the originating market. The integrated response implicitly assumes that 50% of transactions in the originating market separately occur on bid and offer. We find this is a very accurate approximation. To estimate Δ_s^{ES}, we simply sum the median responses in panels (a) and (b) of Figure 27, average them with the sum of median responses in panels (a) and (b) of Figure 28 and integrate over horizon s. Δ_s^{SPY} is analogously estimated with panels (c) and (d) of Figures 27 and 28. Naturally, Δ_s^{ES} is reported in units of SPY shares, as it measures the SPY order flow (negative order flow in the case of bid transactions) at the Nasdaq exchange after an ES transaction at the CME. However, while the natural units of Δ_s^{SPY} would be ES contracts, we have scaled by 500 in order maintain identical units across originating markets. The integrated responses exhibit the same order of magnitude effects that are depicted in the figures: SPY transactions have an effect that is up to 100 times larger than those of ES. Further the first row of panel (c) shows that 72% of the SPY response to ES occurs by 10 ms while the second row reports only 45% of the ES response to SPY occurs in the same time frame.

Panel (c) of Table 2 clearly demonstrates that SPY transactions have a relatively larger impact on the ES market than vice versa. However, given the relative frequency of cross-market arbitrage opportunities at the time of an ES transaction (and the relative paucity of opportunities at the time of a SPY transaction), we expect that the total profit from exploiting these discrepancies will be much more balanced. Panel (d) of Table 2 weights the integrated responses in panel (c) by the average number of transactions per day, and additionally scales by an approximated profit per opportunity. As reported in panel (a) of the table, there are

$16,448,563 \times 3.826 = 62,932,202$ ES contracts traded during our 62-day sample period, resulting in an average $1,015,036$ contracts per day. For the SPY market, we subset just the transactions of 500 shares or more. During our 62-day sample period there were a total of 402,398 such transactions, with a median size of 1002 shares, resulting in approximately 403,202,796 shares during the entire period, or 13,007 lots of 500 shares per day. We use the median to approximate the number of shares traded per day because of the large skew in the SPY volume distribution, attributed to a very small fraction of extremely large trade sizes, which are unlikely to exploit the cross-market mechanisms that we outline in this work. The final step in computing the integrated order flow profit is to approximate the profit per opportunity in each direction. These values are obtained on a daily basis by determining the relative frequency of each profitable opportunity and computing the associated weighted average of profits. This amounts to using the basis-adjusted histograms in Figure 25 as weights for the associated profit opportunities for specific bid and offer differences. Averaging across days we arrive at a profit of 1.367 cents per ES contract event and 0.9396 cents per 500 share SPY event. After scaling the integrated responses of panel (c) by the number of events and profit per event, we see that the integrated response profit is much more balanced – over the 30-millisecond horizon, ES events result in an expected profit of roughly \$183,000 per day and SPY events result in roughly \$171,000. We highlight that these values only represent the expected arbitrage profits between the CME and Nasdaq exchanges; given that Nasdaq accounts for approximately 25% of equities market trading, we anticipate the aggregate size of the trade to be about 4 times larger, per day, than the numbers we report. Our results demonstrate that although the per-event impact of SPY arbitrage opportunities is about two orders of magnitude larger than that of ES opportunities, the aggregate effect is ameliorated by the fact that there are almost two orders of magnitude fewer events in the SPY market that give rise to cross market arbitrage.

While the literature on market information share typically attributes only a small weight to the ETF or cash market (typically around 10%), the foregoing results are striking because they demonstrate the relative information/profit importance of transactions in the ETF market. Further, our results are quite striking in that they resoundingly support the predictions of our model and highlight a previously unknown mechanism of cross-market information flow: relative price increment. The relevance of this mechanism is emphasized by the hundreds of billions of dollars of notional value traded daily in the ES/SPY market, and the trillion dollar size of the futures/spot markets as a whole. From a regulatory perspective, our results suggest that both price increment and contract size are not arbitrary variables in contract and exchange design, but that regulators such as the SEC and CFTC should carefully consider the implications and consequences of contract relationships across equivalent or nearly equivalent assets.

14 Conclusion

The relationship of informed price movements, liquidity changes and order flow are important considerations in fragmented markets that trade equivalent, or very similar assets. These considerations become increasingly more important as markets become faster, and as market makers gain access to technology to intelligently traverse distinct market centers in very short intervals of time. Our work highlights the role of relative contract specifications, especially relative price increments, between two similar assets, in determining the direction of informed order flow between markets. We develop a model which demonstrates that market makers must balance competing needs to quote narrow spreads, in order to promote increased own-market order flow through reduced transactions costs, with the desire to quote wide spreads, in order to create more profitable cross-market arbitrage opportunities. Coupled with a leverage channel, which induces market makers in the futures market to quote wider spreads, our model makes several predictions regarding the direction of informed order flow between markets: when a market maker is passively filled on bid (offer) and the their transacted price is less (more) than the bid (offer) in the other market, they try to lay off some portion of their risk by directing aggressive order flow to the other market, while simultaneously earning an arbitrage profit. Despite preconceived notions of futures/cash market order flow relationships in practice, our model is agnostic of those observed relationships and makes predictions that information should move in both directions across markets, albeit under differing circumstances.

In our empirical study, we show that the contracts for futures and cash instruments on the S&P 500 index (the E-mini and SPY ETF, respectively) are designed so that the minimum price increment of the former is $2.5\times$ larger than that of the latter, resulting in cross-market arbitrage opportunities that typically favor market makers in the E-mini. Indeed, conventional understanding of these two markets dictates that price or liquidity changes in the E-mini almost always precede those of the SPY ETF by approximately the time it takes to transmit messages between the two trading venues – 4 ms. Using new econometric methodology, we show this standard direction of informed order flow is supported in the data: the SPY market typically reacts to the ES market, but only under conditions that our model suggests are favorable for market-maker arbitrage. Likewise, our empirical work shows that the conventional direction of informed order flow is reversed under the same conditions that favor cross-market arbitrage opportunities for SPY market makers. Our econometric method is simple, non-parametric and designed for asynchronous, high-frequency data and in all cases shows that the data strongly support the conclusions of our model.

We live in an era in which exchanges are proliferating, markets are becoming increasingly fragmented, trading technology is quickly changing and in which regulators are pressured to respond to perceived instability. As new exchanges, with differing contract specifications and order matching mechanisms are being added to a complex financial market network, the determinants of informed trading, and its direction within the system, become of vital importance to regulators. Our work demonstrates that a seemingly innocuous control variable, *relative* (not absolute) price increment, can have important implications for cross-market arbitrage, and hence direction of informed order flow. Understanding this channel, as well as others like it, will be necessary for regulators such as the SEC and CFTC and may guide them in further coordinating these distinct markets.

References

Aggarwal, N. and Thomas, S. (2011), "When do stock futures dominate price discovery?" *Working Paper*, 1–37.

Aldrich, E. M., Grundfest., J. A., and Laughlin, G. (2016), "The Flash Crash: A New Deconstruction," *Working Paper*.

Aldrich, E. M., Heckenbach, I., and Laughlin, G. (2015), "The Random Walk of High Frequency Trading," *Working Paper*, 1–39.

Ane, T. and Geman, H. (2000), "Order Flow, Transaction Clock, and Normality of Asset Returns," *The Journal of Finance*, LV, 2259–2284.

Bose, S. (2007), "Contribution of Indian Index Futures to Price Formation in the Stock Market," *Money and Finance*, 3, 39–56.

Brada, J., Ernst, H., and van Tassel, J. (1966), "The Distribution of Stock Price Differences: Gaussian after All?" *Operations Research*, 14, 334–340.

Campbell, B. and Hendry, S. (2007), "Price Discovery in Canadian and U.S. 10-Year Government Bond Markets," *Bank of Canada Working Paper*, 1–47.

Chakravarty, S., Gulen, H., and Mayhew, S. (2004), "Informed trading in stock and option markets," *The Journal of Finance*, 59, 1235–1257.

Choi, H. and Subrahmanyam, A. (1994), "Using Intraday Data to Test for Effects of Index Futures on the Underlying Stock Markets," *The Journal of Futures M*, 14, 293–322.

Chu, Q. C., Hsieh, W.-l. G., and Tse, Y. (1999), "Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs," *International Review of Financial Analysis*, 8, 21–34.

Clark, P. K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

Easley, D., Hara, M. O., and Srinivas, P. S. (1998), "Option Volume and Stock Prices: Evidence on Where Informed Traders Trade," *The Journal of Finance*, LIII, 431–465.

Fleming, J., Ostdiek, B., and Whaley, R. E. (1996), "Trading Costs and the Relative Rates of Price Discovery in Stock, Futures, and Option Markets," *The Journal of Futures Markets*, 16, 353–387.

Grammig, J. G. and Peter, F. J. (2010), "Tell-tale tails: A data driven approach to estimate unique market information shares," *Working Paper*.

Hagströmer, B. and Menkveld, A. J. (2016), "A Network Map of Information Percolation," *Working Paper*.

Harris, L., Sofianos, G., and Shapiro, J. E. (1994), "Program Trading and Intraday Volatility," *The Review of Financial Studies*, 7, 653–685.

Hasbrouck, J. (1995), "One Security, Many Markets: Determining Contributions to Price Discovery," *The Journal of Finance*, 50, 1175–1199.

— (2003), "Intraday Price Formation in U.S. Equity Index Markets," *The Journal of Finance*, LVIII, 2375–2399.

Hendershott, T. and Jones, C. M. (2005), "Island Goes Dark: Transparency, Fragmentation, and Regulation," *Review of Financial Studies*, 18, 743–793.

Kawaller, I. G., Koch, P. D., and Koch, T. W. (1987), "The Temporal Price Relationship between S&P 500 Futures and the S&P 500 Index," *The Journal of Finance*, 42, 1309–1329.

Labuszewski, J. W. and Co, R. (2010), "What Happened on May 6th?" *CME Group Staff Report*.

Laughlin, G., Aguirre, A., and Grundfest, J. (2014), "Information Transmission between Financial Markets in Chicago and New York," *The Financial Review*, 49, 283–312.

MacKinlay, A. C. and Ramaswamy, K. (1988), "Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices," *The Review of Financial Studies*, 1, 137–158.

Mandelbrot, B. and Taylor, H. M. (1967), "On the Distribution of Stock Price Differences," *Operations Research*, 15, 1057–1062.

Roll, R., Schwartz, E., and Subrahmanyam, A. (2014), "Trading activity in the equity market and its contingent claims: An empirical investigation," *Journal of Empirical Finance*, 28, 13–35.

Rosenberg, J. V. and Traub, L. G. (2006), "Price discovery in the foreign currency futures and spot market," *Working Paper*.

Schlusche, B. (2009), "Price Formation in Spot and Futures Markets: Exchange Traded Funds vs. Index Futures," *Journal of Derivatives*, 17, 26–40.

Stephan, J. a. and Whaley, R. E. (1990), "Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets," *The Journal of Finance*, 45, 191.

Stock, J. and Watson, M. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

Stoll, H. R. and Whaley, R. E. (1990), "The Dynamics of Stock Index and Stock Index Futures Returns," *The Journal of Financial and Quantitative Analysis*, 25, 441–468.

Wahab, M. and Lashgari, M. (1993), "Price Dynamics and Error Correction in Stock Index and Stock Index Futures Markets: A Cointegration Approach," *The Journal of Futures*, 13, 711–742.

Yan, B. and Zivot, E. (2007), "The Dynamics of Price Discovery," *Working Paper*.

— (2010), "A structural analysis of price discovery measures," *Journal of Financial Markets*, 13, 1–19.

Yang, J. (2009), "Do Futures Lead Price Discovery in Electronic Foreign Exchange Markets," *The Journal of Futures Markets*, 29, 137–156.

Chapter 3: Incorporation of Dynamic Age-Wage Profile in OLG to model Social Security

This chapter is joint paper with Professors Lilia Maliar of Stanford University and Serguei Maliar of University of Santa Clara. The unpublished article is inserted with their permission.

Abstract

United States population has increasingly become older. The aging is expected to continue as life expectancy increases, more baby boomers reach retirement age and birth rate declines. These demographic changes have had significant impact on the solvency of the Social Security Trust Fund. Social Security Administration projects the fund's insolvency around 2035. Given the imminent insolvency, policies such as increasing eligibility age and expanding the tax base have been proposed. These policies are usually analyzed using many period OLG model with a hump-shaped age-dependent productivity profile. We believe this profile is dynamic as the population itself. We incorporate the dynamic age-dependent profile in an OLG framework in a closer look at insolvency of the Social Security Trust Fund.

Keywords: OLG.

JEL Classification:

15 Introduction

The aging population has hastened the insolvency of the Social Security Trust Fund. The downward trend in birth rate, upward trend in life expectancy, and the retirement of the baby boomers have contributed to the population aging. The aging has increased the proportion of the benefit eligible agents. There also has been a decline in population employment for all age groups except for the group 65 and over. Since Social Security Trust Fund is funded through wage receipts, these changes in the employment participation will adversely affect the sustainability of the fund.

We believe the current labor trends can be attributable to increase in life expectancy and the number of college graduates. Social Security replaces only portion of before retirement income. As agents live longer, it is intuitive that agents will work longer to have enough savings to maintain their consumption needs to their expected death. And as production becomes more skill oriented the less educated/experienced workers will earn less. An optimizing agent will therefore substitute labor between older (more productive) and younger years. In this paper, we will try to address the implications of these labor patterns to the solvency of Social Security Trust Fund.

According to the recent projections by Social Security Administration, the insolvency of the fund is likely to occur between 2030 and 2035 (Figure 29). We believe these estimates do not take into account the labor patterns above. Ceteris paribus the labor patterns will ensure the delay of the insolvency. If older workers never retire then regardless of the composition of the workforce, the Social Security Trust will be sustainable since these workers are not exempt from Social Security tax. And as the older cohorts become more educated than the cohorts before them the employment of 65 and over will continue to increase. However, as emphasis on skill and experience increases, the younger cohorts will reduce labor hours for leisure or more schooling, if we assume wages are proportional productivity. We will show that the substitution of labor between old (65+) and young (18-24) plays a key role in our model.

Although the insolvency can be delayed it is unlikely insolvency is avoidable without policy changes. National Research Council (2012) highlighted the problems of aging population and possible measures to avoid insolvency. The policy measures needed to avoid insolvency will likely require reduction in consumption of the retired or increases in taxes to fund the system. However, both measures are highly unpopular and unlikely to generate enough electoral support. McGrattan, E., Prescott (2015) propose eliminating FICA taxes and broadening the tax base to address the problems of the Social Security Fund. Their study and other analysis of Social Security Trust Fund such as Nishiyama (2013) use a static wage-age profile derived from cross-sectional wage data. We believe the wage-age profile is dynamic and incorporate the dynamism in our model. National Research Council (2012) proposed a possible reduction in payroll tax for the working 65 and over to increase the welfare for this group. We hypothesize this will also delay the insolvency of the Social Security because of the increased efficiency for this age group. We hypothesize lower taxes will increase the employment participation for a group that has been increasing their productivity.

The layout of the paper is as follow. We will provide evidence that these observed labor patterns can be well explained by the rise in higher education. Then using the current estimates of college graduates and future projections of life expectancy we will forecast efficiency weight profile for years 2014 to 2150 (the date of final steady state). The model economy will be simulated using a static and dynamic profiles. Finally, we will detail the implications of dynamic profile on the solvency of the Social Security Trust Fund and the economy.

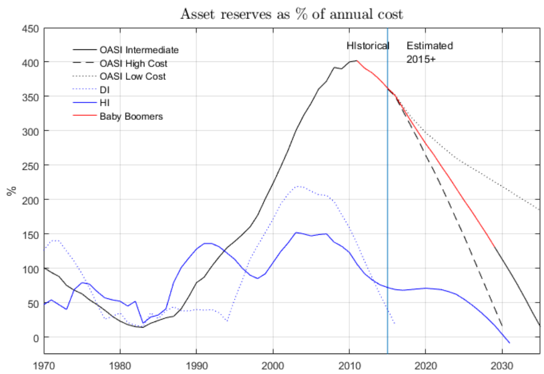

Figure 29: Source: Social Security Bulletin Vol 70 No 3, 2010 and 2015 Annual Report Table IV B4.

16 Empirical

16.1 Social Security

The Social Security Trust Fund is projected to be insolvent in early 2030 (Figure 29). Social Security Trust Fund is funded through a tax of 9.9% up to $118,500 on wage receipts and any interest income from the current surplus. The fund pays benefits to all eligible beneficiaries who elect into the program[11]. The benefits are dependent on past contributions. The number of beneficiaries have been increasing such that the fund's yearly expenditures exceed receipts. The fund's current surplus will allow the Social Security Administration to fulfill its benefits schedule until its depletion, the insolvent date.

The increase in expenditure is easy to account for. The retirees of today are living longer than the retirees before them and the ratio of retirees to workers are at historical highs due to longevity and retirement of baby boomers. Advances in medicine, eradication of many infectious diseases and improved diet have contributed to increases in life expectancy. Figure 30 shows the historical increases in survival probabilities in 5 year increments. The increases have been dramatic. A person 65 years old in 2010 was almost twice as likely to reach the age of 85 than a 65 year old in 1959. These increases are only projected to rise.

Figure 31 shows the projected survival probabilities for each age up to year 2086[12]. According to the projections, in 2085 the probability a 65-year-old reaching the age of 85 will be more than 64% almost a threefold increase from 1959. One measure of sustainability of Social Security is the ratio of 65+ to the 20-64 as shown in Figure 32.

[11] Retirees are able to receive reduced benefits starting at age 62.

[12] Source: SSA Life Tables.

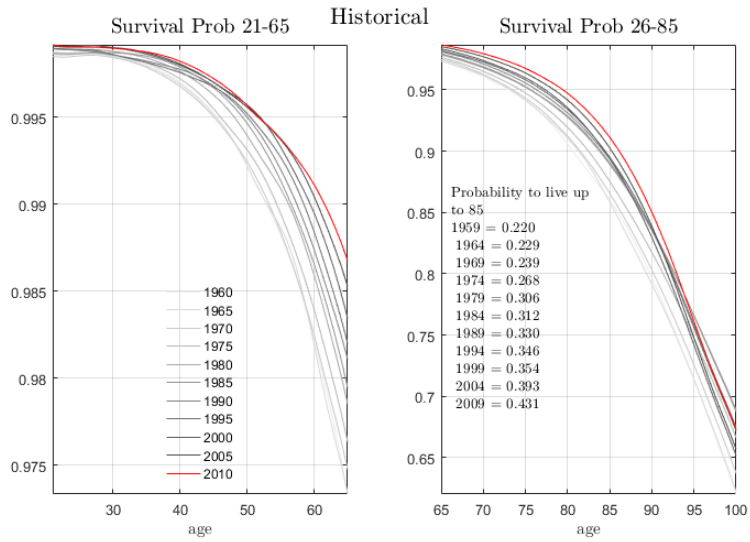

Figure 30: Historical Survival Probabilities. Source: https://www.ssa.gov/oact/NOTES/as116/as116_Tbl_6_1900.html

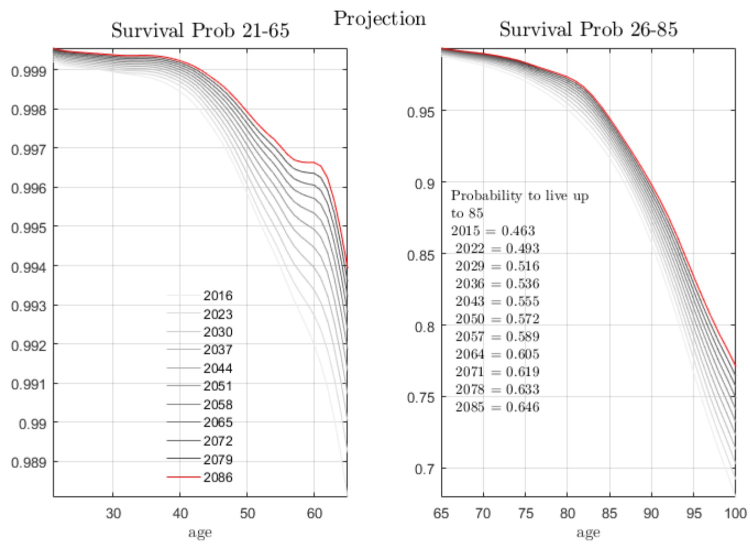

Figure 31: Projected Survival Probabilities by SSA. Source: https://www.ssa.gov/oact/NOTES/as116/as116_Tbl_6_1900.html

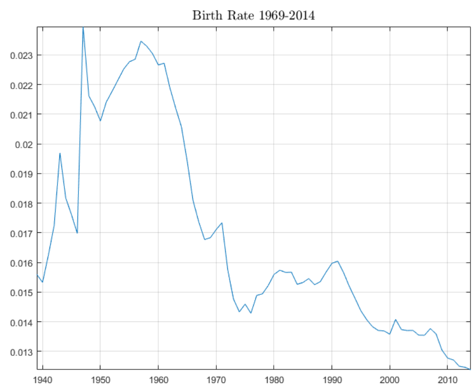

Figure 32: Historical Birth Rate. Source: Census 2010

Figure 33: Age Dependency Ratio

The ratio highlights the Social Security Trust's need for younger cohorts as they are the primary wage earners. However, in the wake of rapidly rising number of beneficiaries, the population has seen a decline in birth rate and employment percentage. These two trends have adversely affected the growth of the contributors to the fund. The decline in birth rate is significant. Number of births were lower in 2014 than in 1965 when the US population was about 200 million compared to 320 million in 2014. Decline in birth rate imply a slower growth of the workforce who are ready to contribute to the Social Security Trust Fund. The recent declining trend in employed population percentage has further slowed the growth of the workforce.

Figure 34 shows the employment percentage for different age groups, with a cross section of different education levels. First population employment percentage has been declining for all ages and education levels as shown by top left in Figure 34. Second employment percentage is approximately the same for ages 25 to 54. The 18-24 and 55-64 have lower employment percentage. The 65 and over have significantly lower employment percentage. For all groups, employment percentage rises for higher levels of education. Jobs associated with higher levels of education are higher paying, making leisure more costly. Manual jobs associated with lower levels of education become more difficult with old age and pay significantly less than skilled jobs. Then cohorts on the verge of retirement will be more likely to delay retirement in professions that are high paying and/or less taxing physically.

Table 4 provides some support for this explanation. The professions with higher pay have higher median age. Professions associated with higher degrees are among the largest increases (profession employment increases in percentage terms) for the oldest age group. Professions associated with manual labor are among the lowest increases with one exception being installation, maintenance and repair occupations. As the percentage of the population with college degrees rises, we expect higher proportion of cohorts to be employed for cohorts older than 24. We will further explore the relationship of education and employment in the following section.

Figure 34: Employment Percentage. Source: ASEC CPS.

Profession	Increase in Occupation			Median	
	Δ All	Δ 65	Δ 20-24	Age	Pay
All	3.30	59.39	-0.90	42.4	32278
Management	9.70	63.89	3.12	47.5	87963
Business and financial operations	17.36	88.06	15.64	43.6	58703
Computer and mathematical	23.71	226.67	28.42	41.1	71640
Architecture and engineering	1.65	78.08	3.82	44.3	68576
Life physicial and social science	-4.34	74.42	-7.06	42.8	55998
Community and social service	7.20	57.27	-5.65	43.9	37550
Legal occupations	15.19	132.26	7.69	46.0	70021
Education, training and library	8.76	112.08	5.08	43.3	42454
Arts, design and entertainment	6.90	79.41	14.50	40.5	41046
Healthcare practitioner and technical	19.67	126.40	21.43	43.7	56237
Health care support	19.14	51.19	27.16	39.1	23997
Protective service	9.48	47.47	6.03	40.8	33833
Food preparation and service	12.02	35.58	16.56	29.4	17501
Building and grounds cleaning	8.78	47.92	-13.28	44.1	21135
Personal care and service	18.35	71.50	28.67	40.0	19332
Sales and related	-2.79	34.48	6.03	39.9	23150
Office and administrative support	-9.01	44.01	-20.10	43.1	29453
Farming Fishing and Forrestry	-2.76	10.00	9.52	37.2	17832
Construction and extraction	-17.83	36.54	-43.77	41.6	37421
Installation maintenance and repair	-2.09	83.50	-8.71	43.0	38130
Production	-13.40	26.36	-8.58	43.0	28754
Transportation and Moving	2.54	56.44	2.28	43.1	26775

Table 4: Changes are percent change from 2004 to 2013. Source: BLS Employed persons by detailed occupation.

16.2 Education

Proportion of 18-24 attending college has been steadily increasing last 50 years (Figure 35). As these students graduate and become older, they replace the cohorts with lower proportion of higher education. Figure 36 shows the mechanical process of cohorts becoming more educated over time. Even if college attendance stagnates (it has actually decreased last few years), we will see an increase in the percentage of college graduates over time as the current old are relatively low in college graduates (in percentage terms) compared to the current young.

To predict future education rates, we regress each cohort's college proportion at time t to fixed effects for ages (21-74)[13], its interaction to cohorts one year younger at time $t-1$ and life expectancy[14]. The model essentially predicts college proportion of cohort at time t, by the college proportion of the one age younger cohort one year earlier. Equation 1 is a reduced form representation of the quantity of college educated cohorts. Our goal here is not inferential and we ignore the usual covariates involved in supply and demand estimation. Our time series model accurately fits the data and the projections seem stationary, which is one of the criteria we required for the projection methods in this section.

Future estimates of college educated cohorts are used in order to construct a time series of future employment percentage and cohort production efficiency. In transition analysis of OLG popularized by Auerbach and Kotlikoff (1987), steady state of the economy is assumed to occur many years ahead i.e. 100+ years. Therefore any forecast in a linear model cannot have a time trend if we assume the variable needs to be at a steady state at some point in the future[15]. However, the variables of interest do have strong time trend. We find by replacing the time trend with life expectancy which we assume converges in 2090 (the date of farthest projections by the Social Security Trust Administration), we can induce stationarity to the projections.

[13] Dummy Variable is for age - 1.

[14] This is defined as the probability of living up to age 85 given one reaches the age of 75. The regression results were robust to different end points of ages used in the definition.

[15] Non-linear estimation techniques can incorporate a time trend and still converge.

$$C_{i,t} = \alpha + \beta_1 \mathbb{1}_{(i-1=21)} + \beta_2 \mathbb{1}_{(i-1=22)} + ... + \beta_{54} \mathbb{1}_{(i-1=74)} + \beta_{55} \mathbb{1}_{(i-1=20)} C_{20,t-1}$$
$$+\beta_{56} \mathbb{1}_{(i-1=21)} C_{21,t-1} + ... + \beta_{109} \mathbb{1}_{(i-1=74)} C_{74,t-1} + e_{i,t} \quad (76)$$

for $i = 21, 22, ... 75$ and $t = 1977$ to 2013[16].

Dependent: ratio of college graduates to cohort size of age i at time t							
Variable	Estimate	SE	tStat	Variable	Estimate	SE	tStat
constant	0.0030	0.0029	1.0214				
				$\mathbb{1}(age-1=20)C_{20,t-1}$	3.0480	0.0038	1.7813
$\mathbb{1}(age-1=21)$	0.0654	0.0046	14.0849	$\mathbb{1}(age-1=21)C_{21,t-1}$	1.8160	0.2239	8.1121
$\mathbb{1}(age-1=22)$	-0.0579	0.0124	-4.6650	$\mathbb{1}(age-1=22)C_{22,t-1}$	2.4122	0.1244	19.3970
$\mathbb{1}(age-1=23)$	-0.0360	0.0066	-5.4302	$\mathbb{1}(age-1=23)C_{23,t-1}$	1.4131	0.0335	42.2029
$\mathbb{1}(age-1=24)$	0.0037	0.0059	0.6236	$\mathbb{1}(age-1=24)C_{24,t-1}$	1.0546	0.0236	44.7422
$\mathbb{1}(age-1=25)$	-0.0125	0.0061	-2.0531	$\mathbb{1}(age-1=25)C_{25,t-1}$	1.1169	0.0230	48.5146
$\mathbb{1}(age-1=26)$	-0.0189	0.0061	-3.1124	$\mathbb{1}(age-1=26)C_{26,t-1}$	1.1176	0.0216	51.7081
$\mathbb{1}(age-1=27)$	-0.0074	0.0060	-1.2424	$\mathbb{1}(age-1=27)C_{27,t-1}$	1.0336	0.0203	50.8596
$\mathbb{1}(age-1=28)$	-0.0234	0.0059	-3.9742	$\mathbb{1}(age-1=28)C_{28,t-1}$	1.1391	0.0199	57.2824
$\mathbb{1}(age-1=29)$	0.0246	0.0056	4.3460	$\mathbb{1}(age-1=29)C_{29,t-1}$	0.9317	0.0179	52.0824
$\mathbb{1}(age-1=30)$	-0.0005	0.0062	-0.0797	$\mathbb{1}(age-1=30)C_{30,t-1}$	0.9971	0.0197	50.7104
$\mathbb{1}(age-1=31)$	-0.0195	0.0063	-3.1167	$\mathbb{1}(age-1=31)C_{31,t-1}$	1.0798	0.0202	53.5313
$\mathbb{1}(age-1=32)$	-0.0097	0.0061	-1.6045	$\mathbb{1}(age-1=32)C_{32,t-1}$	1.0266	0.0192	53.5601
$\mathbb{1}(age-1=33)$	0.0072	0.0060	1.1999	$\mathbb{1}(age-1=33)C_{33,t-1}$	0.9849	0.0191	51.4700
$\mathbb{1}(age-1=34)$	-0.0254	0.0061	-4.1796	$\mathbb{1}(age-1=34)C_{34,t-1}$	1.1101	0.0193	57.4387
$\mathbb{1}(age-1=35)$	-0.0027	0.0057	-0.4770	$\mathbb{1}(age-1=35)C_{35,t-1}$	1.0003	0.0174	57.3642
$\mathbb{1}(age-1=36)$	-0.0175	0.0056	-3.1399	$\mathbb{1}(age-1=36)C_{36,t-1}$	1.0631	0.0173	61.5003
$\mathbb{1}(age-1=37)$	0.0073	0.0053	1.3659	$\mathbb{1}(age-1=37)C_{37,t-1}$	0.9713	0.0163	59.6754
$\mathbb{1}(age-1=38)$	-0.0192	0.0053	-3.5828	$\mathbb{1}(age-1=38)C_{38,t-1}$	1.0679	0.0166	64.3028
$\mathbb{1}(age-1=39)$	0.0141	0.0051	2.7928	$\mathbb{1}(age-1=39)C_{39,t-1}$	0.9780	0.0155	63.2523
$\mathbb{1}(age-1=40)$	0.0170	0.0051	3.3376	$\mathbb{1}(age-1=40)C_{40,t-1}$	0.9150	0.0153	59.6605
$\mathbb{1}(age-1=41)$	0.0124	0.0052	2.3686	$\mathbb{1}(age-1=41)C_{41,t-1}$	0.9429	0.0164	57.3363
$\mathbb{1}(age-1=42)$	-0.0140	0.0053	-2.6564	$\mathbb{1}(age-1=42)C_{42,t-1}$	1.0521	0.0170	61.7688
$\mathbb{1}(age-1=43)$	0.0080	0.0050	1.6156	$\mathbb{1}(age-1=43)C_{43,t-1}$	0.9587	0.0156	61.3315
$\mathbb{1}(age-1=44)$	0.0074	0.0049	1.5122	$\mathbb{1}(age-1=44)C_{44,t-1}$	0.9890	0.0157	62.9442
$\mathbb{1}(age-1=45)$	-0.0187	0.0049	-3.8024	$\mathbb{1}(age-1=45)C_{45,t-1}$	1.0493	0.0155	67.7918
$\mathbb{1}(age-1=46)$	0.0012	0.0045	0.2566	$\mathbb{1}(age-1=46)C_{46,t-1}$	0.9891	0.0140	70.4169
$\mathbb{1}(age-1=47)$	0.0136	0.0044	3.0613	$\mathbb{1}(age-1=47)C_{47,t-1}$	0.9326	0.0136	68.5102
$\mathbb{1}(age-1=48)$	-0.0043	0.0045	-0.9722	$\mathbb{1}(age-1=48)C_{48,t-1}$	1.0123	0.0141	71.8614
$\mathbb{1}(age-1=49)$	0.0003	0.0043	0.0610	$\mathbb{1}(age-1=49)C_{49,t-1}$	1.0223	0.0135	75.9646
$\mathbb{1}(age-1=50)$	-0.0005	0.0042	-0.1099	$\mathbb{1}(age-1=50)C_{50,t-1}$	0.9698	0.0127	76.3320
$\mathbb{1}(age-1=51)$	0.0057	0.0042	1.3738	$\mathbb{1}(age-1=51)C_{51,t-1}$	0.9485	0.0127	74.6761
$\mathbb{1}(age-1=52)$	-0.0054	0.0041	-1.3096	$\mathbb{1}(age-1=52)C_{52,t-1}$	1.0217	0.0129	79.2462
$\mathbb{1}(age-1=53)$	0.0133	0.0040	3.3262	$\mathbb{1}(age-1=53)C_{53,t-1}$	0.9495	0.0123	77.3463
$\mathbb{1}(age-1=54)$	-0.0073	0.0041	-1.8028	$\mathbb{1}(age-1=54)C_{54,t-1}$	1.0301	0.0126	82.0132
$\mathbb{1}(age-1=55)$	-0.0044	0.0039	-1.1096	$\mathbb{1}(age-1=55)C_{55,t-1}$	0.9875	0.0119	82.9685
$\mathbb{1}(age-1=56)$	-0.0133	0.0039	-3.4358	$\mathbb{1}(age-1=56)C_{56,t-1}$	1.0609	0.0118	89.7314
$\mathbb{1}(age-1=57)$	0.0032	0.0037	0.8449	$\mathbb{1}(age-1=57)C_{57,t-1}$	0.9646	0.0110	87.4295
$\mathbb{1}(age-1=58)$	0.0001	0.0037	0.0265	$\mathbb{1}(age-1=58)C_{58,t-1}$	0.9806	0.0115	85.2953
$\mathbb{1}(age-1=59)$	-0.0144	0.0037	-3.8532	$\mathbb{1}(age-1=59)C_{59,t-1}$	1.1046	0.0118	93.5721
$\mathbb{1}(age-1=60)$	-0.0082	0.0036	-2.2572	$\mathbb{1}(age-1=60)C_{60,t-1}$	1.0269	0.0108	95.2062
$\mathbb{1}(age-1=61)$	0.0076	0.0036	2.1185	$\mathbb{1}(age-1=61)C_{61,t-1}$	0.9309	0.0107	86.6512
$\mathbb{1}(age-1=62)$	-0.0083	0.0036	-2.2685	$\mathbb{1}(age-1=62)C_{62,t-1}$	1.0369	0.0117	88.2525
$\mathbb{1}(age-1=63)$	-0.0035	0.0036	-0.9692	$\mathbb{1}(age-1=63)C_{63,t-1}$	1.0089	0.0116	86.6551
$\mathbb{1}(age-1=64)$	0.0055	0.0036	1.5293	$\mathbb{1}(age-1=64)C_{64,t-1}$	0.9880	0.0119	83.1922
$\mathbb{1}(age-1=65)$	-0.0088	0.0036	-2.4081	$\mathbb{1}(age-1=65)C_{65,t-1}$	0.9909	0.0124	79.8082
$\mathbb{1}(age-1=66)$	-0.0035	0.0036	-0.9689	$\mathbb{1}(age-1=66)C_{66,t-1}$	1.0082	0.0132	76.5209
$\mathbb{1}(age-1=67)$	0.0039	0.0036	1.0744	$\mathbb{1}(age-1=67)C_{67,t-1}$	0.9579	0.0139	68.7877
$\mathbb{1}(age-1=68)$	-0.0010	0.0037	-0.2547	$\mathbb{1}(age-1=68)C_{68,t-1}$	0.9957	0.0154	64.5789
$\mathbb{1}(age-1=69)$	-0.0053	0.0038	-1.4110	$\mathbb{1}(age-1=69)C_{69,t-1}$	1.0261	0.0162	63.1752
$\mathbb{1}(age-1=70)$	-0.0083	0.0038	-2.2081	$\mathbb{1}(age-1=70)C_{70,t-1}$	1.0442	0.0164	63.7795
$\mathbb{1}(age-1=71)$	-0.0059	0.0037	-1.5967	$\mathbb{1}(age-1=71)C_{71,t-1}$	1.0093	0.0163	62.0104
$\mathbb{1}(age-1=72)$	0.0008	0.0037	0.2145	$\mathbb{1}(age-1=72)C_{72,t-1}$	0.9587	0.0167	57.3361
$\mathbb{1}(age-1=73)$	0.0021	0.0037	0.5648	$\mathbb{1}(age-1=73)C_{73,t-1}$	0.9706	0.0182	53.2632
$\mathbb{1}(age-1=74)$	0.0067	0.0038	1.7813	$\mathbb{1}(age-1=74)C_{74,t-1}$	0.9564	0.0191	50.1636

Table 5: Regression results for eqn 1. R^2=.997. Fstat=5.13e3. p-value=0.

[16] The model regression results assume homoscedasticity, where regression errors $e_{i,t}$ are independent on i and t.

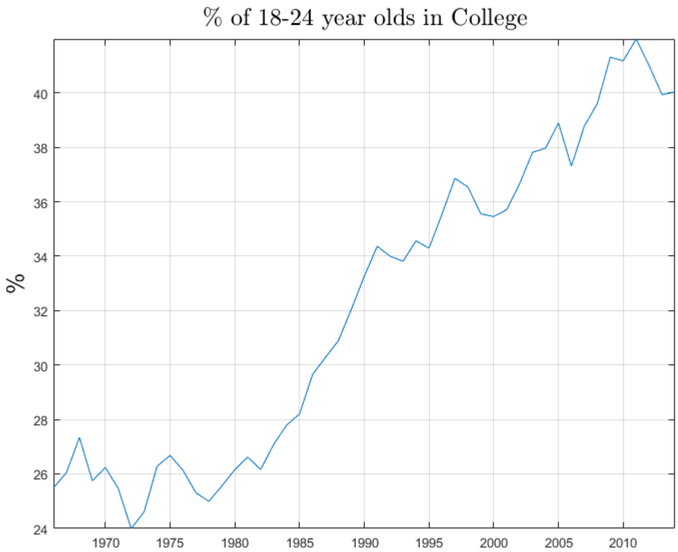

Figure 35: Source: http://www.census.gov/hhes/school/data/cps/historical/

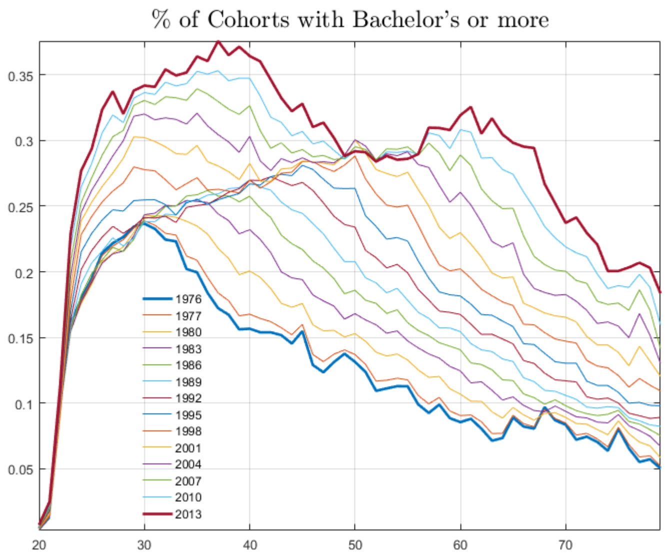

Figure 36: Historical. Source: ASEC CPS

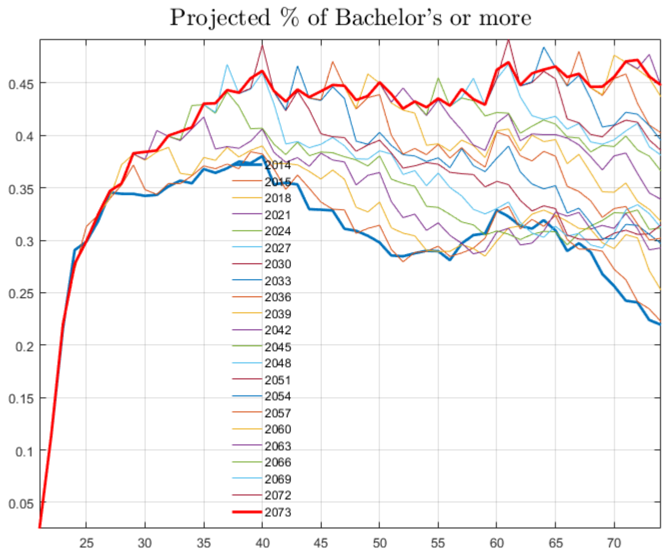

Figure 37: Projected

Figure 37 shows the predictions of the model up to year 2073. The model predicts almost a homogeneous ratio of college graduates after the age of 40 as would be the case if college graduate ratio stabilizes and the lower percentage age groups are replaced over time. The oldest age group will have the largest increase in college graduate proportions. We will see what this implies in employment rates in the following section.

16.3 Employment Percentage

Extensive margins account for most of the fluctuations in labor movement as highlighted by Heckman (1984). Our model is not able to capture endogenously both intensive and extensive margins of labor. We take the view, extensive margin should be accounted for in the model. However, in our model paradigm, agents' labor choices do not generate interesting extensive labor choices without adding an exogenous employment participation[17]. We will assume employment participation is an exogenous process that is a function of time and education level of the cohorts. Our regression model for employment percentage is similar to the model for education (equation 1).

[17] Agents will choose to work when they are most productive, implying they will all work when they are old.

$$E_{i,t} = \alpha + \beta_1 \mathbb{1}_{(i=21)} + \beta_2 \mathbb{1}_{(i=22)} + ... + \beta_{55} \mathbb{1}_{(i=75)} + \beta_{56} \mathbb{1}_{(i=20)} C_{20,t} + \beta_{57} \mathbb{1}_{(i=21)} C_{21,t} + ... + \beta_{111} \mathbb{1}_{(i=75)} C_{75,t} + \beta_{112} LE_t + e_{i,t} \tag{77}$$

for i = 20, 22, ... 75 and t = 1976 to 2013.

The regression results are displayed in Table 6. Life expectancy has a negative coefficient. Although this may be counter intuitive as longer life necessitates more saving therefore longer employment, note there is no time trend in this regression. Life expectancy is pseudo time trend in this regression, as it is a monotonically increasing. Loss in manufacturing jobs to other countries and to machines can explain the negative coefficient. For the youngest cohorts, the coefficients to college proportion is strongly negative. This implies the demand for young educated workers are less than for older workers with more experience. The linear model predicts most of the age groups will be employed at a higher percentage as the proportion of college educated rises (Figure 39).

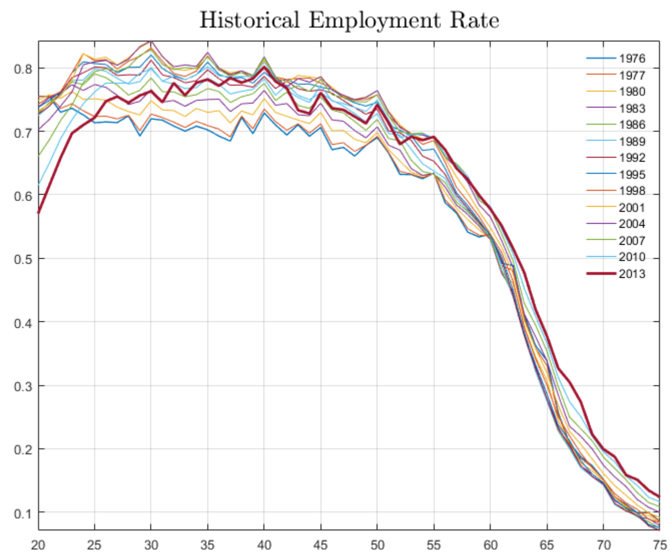

Figure 38: Historical Source: ASEC CPS

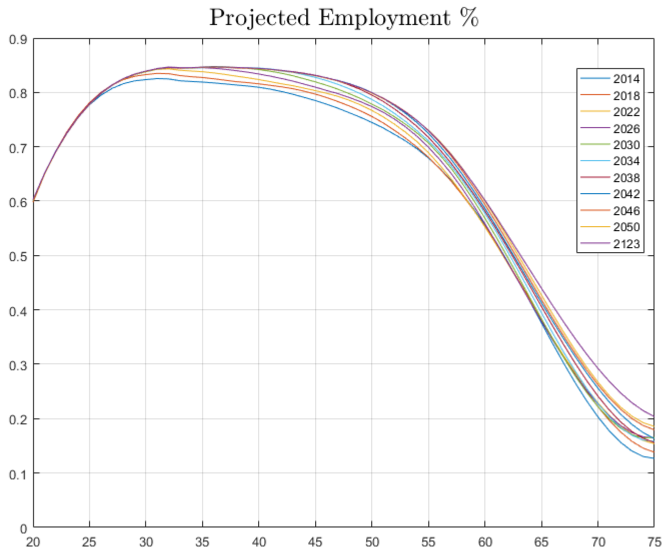

Figure 39: Projected

Dependent: Employment percentage of cohort i at time t							
Variable	Estimate	SE	tStat	Variable	Estimate	SE	tStat
constant	0.9332	0.0243	38.4439				
Life Expectancy	-0.1940	0.0435	-4.4635				
				$\mathbb{1}(age=20)C_{20,t}$	-27.7719	0.0224	-7.0068
$\mathbb{1}(age=21)$	0.769	0.0154	5.005	$\mathbb{1}(age=21)C_{21,t}$	-7.5795	1.3164	-5.7578
$\mathbb{1}(age=22)$	0.0178	0.0270	0.6607	$\mathbb{1}(age=22)C_{22,t}$	-3.5825	0.7315	-4.8977
$\mathbb{1}(age=23)$	0.2592	0.0723	3.5848	$\mathbb{1}(age=23)C_{23,t}$	-0.4391	0.2084	-2.1066
$\mathbb{1}(age=24)$	0.0112	0.0396	0.2834	$\mathbb{1}(age=24)C_{24,t}$	0.3370	0.1494	2.2564
$\mathbb{1}(age=25)$	-0.1261	0.0354	-3.5659	$\mathbb{1}(age=25)C_{25,t}$	0.3527	0.1461	2.4146
$\mathbb{1}(age=26)$	-0.1295	0.0367	-3.5258	$\mathbb{1}(age=26)C_{26,t}$	0.3972	0.1349	2.9432
$\mathbb{1}(age=27)$	-0.1437	0.0362	-3.9676	$\mathbb{1}(age=27)C_{27,t}$	0.4762	0.1271	3.7476
$\mathbb{1}(age=28)$	-0.1766	0.0356	-4.9586	$\mathbb{1}(age=28)C_{28,t}$	0.6319	0.1270	4.9767
$\mathbb{1}(age=29)$	-0.2168	0.0358	-6.0616	$\mathbb{1}(age=29)C_{29,t}$	0.7516	0.1143	6.5765
$\mathbb{1}(age=30)$	-0.2513	0.0342	-7.3463	$\mathbb{1}(age=30)C_{30,t}$	0.6297	0.1252	5.0300
$\mathbb{1}(age=31)$	-0.2062	0.0374	-5.5133	$\mathbb{1}(age=31)C_{31,t}$	0.5945	0.1283	4.6356
$\mathbb{1}(age=32)$	-0.2156	0.0380	-5.6757	$\mathbb{1}(age=32)C_{32,t}$	0.5813	0.1208	4.8107
$\mathbb{1}(age=33)$	-0.2181	0.0364	-5.9887	$\mathbb{1}(age=33)C_{33,t}$	0.6154	0.1207	5.1001
$\mathbb{1}(age=34)$	-0.2317	0.0361	-6.4176	$\mathbb{1}(age=34)C_{34,t}$	0.6538	0.1221	5.3563
$\mathbb{1}(age=35)$	-0.2375	0.0366	-6.4892	$\mathbb{1}(age=35)C_{35,t}$	0.6990	0.1101	6.3478
$\mathbb{1}(age=36)$	-0.2402	0.0341	-7.0402	$\mathbb{1}(age=36)C_{36,t}$	0.6040	0.1090	5.5429
$\mathbb{1}(age=37)$	-0.2239	0.0335	-6.6925	$\mathbb{1}(age=37)C_{37,t}$	0.6617	0.1013	6.5308
$\mathbb{1}(age=38)$	-0.2483	0.0315	-7.8809	$\mathbb{1}(age=38)C_{38,t}$	0.5426	0.1039	5.2206
$\mathbb{1}(age=39)$	-0.2068	0.0318	-6.4939	$\mathbb{1}(age=39)C_{39,t}$	0.5824	0.0964	6.0436
$\mathbb{1}(age=40)$	-0.2198	0.0299	-7.3425	$\mathbb{1}(age=40)C_{40,t}$	0.5923	0.0969	6.1155
$\mathbb{1}(age=41)$	-0.2033	0.0305	-6.6761	$\mathbb{1}(age=41)C_{41,t}$	0.5651	0.1025	5.5115
$\mathbb{1}(age=42)$	-0.2150	0.0310	-6.9290	$\mathbb{1}(age=42)C_{42,t}$	0.5759	0.1065	5.4085
$\mathbb{1}(age=43)$	-0.2186	0.0315	-6.9452	$\mathbb{1}(age=43)C_{43,t}$	0.5199	0.0995	5.2231
$\mathbb{1}(age=44)$	-0.2083	0.0298	-6.9894	$\mathbb{1}(age=44)C_{44,t}$	0.5694	0.1003	5.6778
$\mathbb{1}(age=45)$	-0.2235	0.0296	-7.5431	$\mathbb{1}(age=45)C_{45,t}$	0.5237	0.0987	5.3062
$\mathbb{1}(age=46)$	-0.2024	0.0296	-6.8492	$\mathbb{1}(age=46)C_{46,t}$	0.5872	0.0902	6.5091
$\mathbb{1}(age=47)$	-0.2321	0.0274	-8.4777	$\mathbb{1}(age=47)C_{47,t}$	0.4950	0.0874	5.6608
$\mathbb{1}(age=48)$	-0.2145	0.0266	-8.0576	$\mathbb{1}(age=48)C_{48,t}$	0.5698	0.0908	6.2749
$\mathbb{1}(age=49)$	-0.2396	0.0269	-8.9042	$\mathbb{1}(age=49)C_{49,t}$	0.5705	0.0871	6.5471
$\mathbb{1}(age=50)$	-0.2426	0.0261	-9.2812	$\mathbb{1}(age=50)C_{50,t}$	0.5206	0.0824	6.3194
$\mathbb{1}(age=51)$	-0.2197	0.0256	-8.5846	$\mathbb{1}(age=51)C_{51,t}$	0.4851	0.0824	5.8864
$\mathbb{1}(age=52)$	-0.2382	0.0250	-9.5151	$\mathbb{1}(age=52)C_{52,t}$	0.4269	0.0837	5.1034
$\mathbb{1}(age=53)$	-0.2381	0.0248	-9.6072	$\mathbb{1}(age=53)C_{53,t}$	0.4894	0.0798	6.1352
$\mathbb{1}(age=54)$	-0.2629	0.0241	-10.9261	$\mathbb{1}(age=54)C_{54,t}$	0.5600	0.0817	6.8547
$\mathbb{1}(age=55)$	-0.2895	0.0244	-11.8798	$\mathbb{1}(age=55)C_{55,t}$	0.5062	0.0774	6.5395
$\mathbb{1}(age=56)$	-0.2778	0.0236	-11.7515	$\mathbb{1}(age=56)C_{56,t}$	0.4870	0.0766	6.3539
$\mathbb{1}(age=57)$	-0.2957	0.0231	-12.8072	$\mathbb{1}(age=57)C_{57,t}$	0.4945	0.0711	6.9526
$\mathbb{1}(age=58)$	-0.3222	0.0223	-14.4774	$\mathbb{1}(age=58)C_{58,t}$	0.5119	0.0737	6.9431
$\mathbb{1}(age=59)$	-0.3425	0.0223	-15.3804	$\mathbb{1}(age=59)C_{59,t}$	0.4093	0.0753	5.4348
$\mathbb{1}(age=60)$	-0.3436	0.0222	-15.4986	$\mathbb{1}(age=60)C_{60,t}$	0.3609	0.0689	5.2373
$\mathbb{1}(age=61)$	-0.3537	0.0216	-16.3918	$\mathbb{1}(age=61)C_{61,t}$	0.4400	0.0681	6.4593
$\mathbb{1}(age=62)$	-0.4072	0.0213	-19.1347	$\mathbb{1}(age=62)C_{62,t}$	0.3962	0.0745	5.3178
$\mathbb{1}(age=63)$	-0.4334	0.0216	-20.0920	$\mathbb{1}(age=63)C_{63,t}$	0.4417	0.0729	6.0548
$\mathbb{1}(age=64)$	-0.5053	0.0212	-23.7929	$\mathbb{1}(age=64)C_{64,t}$	0.4966	0.0747	6.6511
$\mathbb{1}(age=65)$	-0.5602	0.0212	-26.4423	$\mathbb{1}(age=65)C_{65,t}$	0.5296	0.0781	6.7787
$\mathbb{1}(age=66)$	-0.6074	0.0215	-28.2350	$\mathbb{1}(age=66)C_{66,t}$	0.5931	0.0815	7.2745
$\mathbb{1}(age=67)$	-0.6644	0.0213	-31.2320	$\mathbb{1}(age=67)C_{67,t}$	0.5971	0.0851	7.0193
$\mathbb{1}(age=68)$	-0.6970	0.0213	-32.6471	$\mathbb{1}(age=68)C_{68,t}$	0.7160	0.0952	7.5220
$\mathbb{1}(age=69)$	-0.7383	0.0219	-33.6521	$\mathbb{1}(age=69)C_{69,t}$	0.6067	0.1009	6.0104
$\mathbb{1}(age=70)$	-0.7400	0.0222	-33.3013	$\mathbb{1}(age=70)C_{70,t}$	0.5341	0.1031	5.1811
$\mathbb{1}(age=71)$	-0.7464	0.0222	-33.6077	$\mathbb{1}(age=71)C_{71,t}$	0.6258	0.1017	6.1553
$\mathbb{1}(age=72)$	-0.7808	0.0219	-35.6630	$\mathbb{1}(age=72)C_{72,t}$	0.5768	0.1048	5.5054
$\mathbb{1}(age=73)$	-0.7853	0.0218	-36.0474	$\mathbb{1}(age=73)C_{73,t}$	0.5864	0.1135	5.1670
$\mathbb{1}(age=74)$	-0.7961	0.0221	-36.0952	$\mathbb{1}(age=74)C_{74,t}$	0.4981	0.1208	4.1230
$\mathbb{1}(age=75)$	-0.7949	0.0224	-35.4982	$\mathbb{1}(age=75)C_{75,t}$	0.5927	0.1291	4.5909

Table 6: Regression results for eqn 2. R^2=.986. Fstat=1.27e3. p-value=0.

16.4 Efficiency

In an OLG model, a hump shaped age-productivity profile is used to better match the age-dependent labor profile. The age dependent productivity is multiplied to each cohort's wage, changing the labor choice to reflect rising/decreasing opportunity cost of leisure. Using the proportional relationship between productivity and wages in a Cobb Douglas production, Hansen (1993) calculates efficiency unit by normalizing hourly wages of various subgroups by the hourly wage of the population for the period 1979 to 1987. Using Hansen's methodology we calculate the age-dependent productivity profile from 1976 to 2013[18]. Figure 40 shows the evolution of the efficiency weights for the period[19].

We highlight three features of Figure 40. First the productivity of the 65+ have increased over 50% since 1976. Second the peak of the highest productivity occurs at later ages with time. Third the younger age group's productivity relative to the peak is smaller today than it was in 1976. We believe the age-dependent profile has sufficiently changed that Hansen's 1993 curve is no longer accurate. The implications of the new profile match the current labor patterns. Higher proportion of old workers delays full retirement as higher wages increase the opportunity cost of retirement. Older workers if they are physically capable have many years of valuable work experience which contribute to their higher productivity. Lemieux (2006) shows wages are increasing in experience controlling for level of education. The lower productivity for the younger workforce makes schooling or pure unemployment more attractive.

[18] We use a different data set. He uses BLS aggregation of the CPS. We use ASEC of the CPS, in order to create subgroups by education.

[19] Selected years for visual aid.

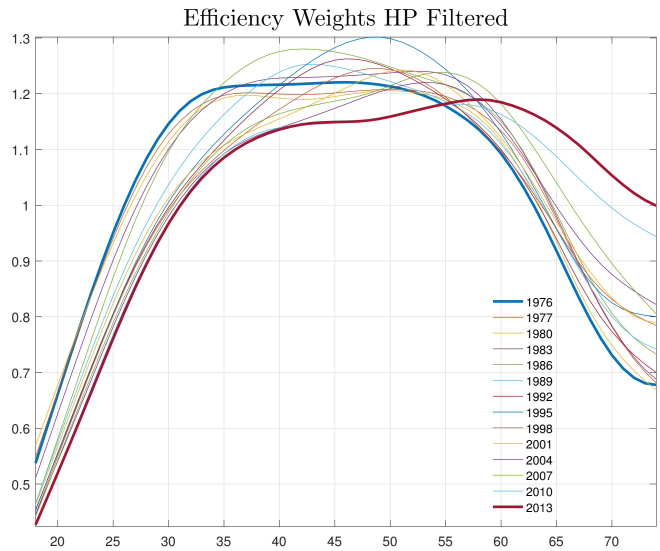

Figure 40: Efficiency Weights 1976-2013 Using Hansen's methodology. Data: ASEC CPS.

In order to forecast age dependent productivity, we use a different specification than Hansen's efficiency weight. We project log hourly wages and normalize them. We do this for two reasons. First Mincer equation, which links log wages to education has a large literature supporting its form. Second when Hansen's efficiency weights are used as the dependent variable, the coefficient for Life Expectancy is negative. This implies the efficiency for all age groups will be declining, which is counter intuitive. The coefficient for life expectancy is positive for the regression specification in equation 3, which provides a productivity profile that is not declining over time.

$$\begin{aligned}\log w_{i,t} = \alpha + \beta_1 \mathbb{1}_{(i=21)} + \beta_2 \mathbb{1}_{(i=22)} + ... + \beta_{55} \mathbb{1}_{(i=75)} + \beta_{56} \mathbb{1}_{(i=20)} C_{20,t} \\ + \beta_{57} \mathbb{1}_{(i=21)} C_{21,t} + ... + \beta_{111} \mathbb{1}_{(i=75)} C_{75,t} + \beta_{112} LE_t + e_{i,t}\end{aligned} \tag{78}$$

for i = 20, 22, ... 75 and t = 1976 to 2013.

	Dependent: log hourly wage of age i at time t						
Variable	Estimate	SE	tStat	Variable	Estimate	SE	tStat
constant	1.6692	0.0215	77.7918				
Life Expectancy	0.5789	0.0384	15.0697				
				$\mathbb{1}(age = 20)C_{20,t}$	24.6504	0.0198	7.0358
$\mathbb{1}(age = 21)$	1.6692	0.0215	77.7918	$\mathbb{1}(age = 21)C_{21,t}$	7.5648	1.1636	6.5011
$\mathbb{1}(age = 22)$	0.0735	0.0238	3.0853	$\mathbb{1}(age = 22)C_{22,t}$	2.9787	0.6466	4.6068
$\mathbb{1}(age = 23)$	-0.0138	0.0639	-0.2162	$\mathbb{1}(age = 23)C_{23,t}$	1.1907	0.1842	6.4631
$\mathbb{1}(age = 24)$	0.1379	0.0350	3.9352	$\mathbb{1}(age = 24)C_{24,t}$	0.8869	0.1320	6.7171
$\mathbb{1}(age = 25)$	0.2264	0.0313	7.2431	$\mathbb{1}(age = 25)C_{25,t}$	0.9401	0.1291	7.2801
$\mathbb{1}(age = 26)$	0.2615	0.0325	8.0578	$\mathbb{1}(age = 26)C_{26,t}$	0.9307	0.1193	7.8028
$\mathbb{1}(age = 27)$	0.3047	0.0320	9.5173	$\mathbb{1}(age = 27)C_{27,t}$	0.9348	0.1123	8.3220
$\mathbb{1}(age = 28)$	0.3430	0.0315	10.8942	$\mathbb{1}(age = 28)C_{28,t}$	1.0639	0.1122	9.4793
$\mathbb{1}(age = 29)$	0.3511	0.0316	11.1081	$\mathbb{1}(age = 29)C_{29,t}$	1.0091	0.1010	9.9890
$\mathbb{1}(age = 30)$	0.3895	0.0302	12.8823	$\mathbb{1}(age = 30)C_{30,t}$	1.1407	0.1107	10.3080
$\mathbb{1}(age = 31)$	0.3785	0.0331	11.4495	$\mathbb{1}(age = 31)C_{31,t}$	1.2084	0.1134	10.6591
$\mathbb{1}(age = 32)$	0.3884	0.0336	11.5679	$\mathbb{1}(age = 32)C_{32,t}$	1.1502	0.1068	10.7690
$\mathbb{1}(age = 33)$	0.4252	0.0322	13.2090	$\mathbb{1}(age = 33)C_{33,t}$	1.1812	0.1067	11.0739
$\mathbb{1}(age = 34)$	0.4397	0.0319	13.7797	$\mathbb{1}(age = 34)C_{34,t}$	1.2221	0.1079	11.3275
$\mathbb{1}(age = 35)$	0.4428	0.0324	13.6849	$\mathbb{1}(age = 35)C_{35,t}$	1.1540	0.0973	11.8555
$\mathbb{1}(age = 36)$	0.4719	0.0302	15.6488	$\mathbb{1}(age = 36)C_{36,t}$	1.1494	0.0963	11.9319
$\mathbb{1}(age = 37)$	0.4903	0.0296	16.5787	$\mathbb{1}(age = 37)C_{37,t}$	1.1004	0.0896	12.2872
$\mathbb{1}(age = 38)$	0.5165	0.0279	18.5427	$\mathbb{1}(age = 38)C_{38,t}$	1.1738	0.0919	12.7765
$\mathbb{1}(age = 39)$	0.5085	0.0282	18.0622	$\mathbb{1}(age = 39)C_{39,t}$	1.1176	0.0852	13.1211
$\mathbb{1}(age = 40)$	0.5346	0.0265	20.2018	$\mathbb{1}(age = 40)C_{40,t}$	1.1971	0.0856	13.9833
$\mathbb{1}(age = 41)$	0.5129	0.0269	19.0526	$\mathbb{1}(age = 41)C_{41,t}$	1.2817	0.0906	14.1428
$\mathbb{1}(age = 42)$	0.5055	0.0274	18.4348	$\mathbb{1}(age = 42)C_{42,t}$	1.4102	0.0941	14.9831
$\mathbb{1}(age = 43)$	0.4830	0.0278	17.3561	$\mathbb{1}(age = 43)C_{43,t}$	1.4268	0.0880	16.2146
$\mathbb{1}(age = 44)$	0.4866	0.0263	18.4714	$\mathbb{1}(age = 44)C_{44,t}$	1.4958	0.0886	16.8731
$\mathbb{1}(age = 45)$	0.4789	0.0262	18.2828	$\mathbb{1}(age = 45)C_{45,t}$	1.5120	0.0872	17.3301
$\mathbb{1}(age = 46)$	0.4730	0.0261	18.1067	$\mathbb{1}(age = 46)C_{46,t}$	1.4376	0.0797	18.0275
$\mathbb{1}(age = 47)$	0.5065	0.0242	20.9293	$\mathbb{1}(age = 47)C_{47,t}$	1.4235	0.0773	18.4157
$\mathbb{1}(age = 48)$	0.5169	0.0235	21.9670	$\mathbb{1}(age = 48)C_{48,t}$	1.5247	0.0803	18.9942
$\mathbb{1}(age = 49)$	0.5007	0.0238	21.0507	$\mathbb{1}(age = 49)C_{49,t}$	1.5132	0.0770	19.6471
$\mathbb{1}(age = 50)$	0.5088	0.0231	22.0242	$\mathbb{1}(age = 50)C_{50,t}$	1.4642	0.0728	20.1080
$\mathbb{1}(age = 51)$	0.5147	0.0226	22.7572	$\mathbb{1}(age = 51)C_{51,t}$	1.4829	0.0729	20.3550
$\mathbb{1}(age = 52)$	0.5232	0.0221	23.6434	$\mathbb{1}(age = 52)C_{52,t}$	1.5284	0.0739	20.6690
$\mathbb{1}(age = 53)$	0.5234	0.0219	23.8943	$\mathbb{1}(age = 53)C_{53,t}$	1.4727	0.0705	20.8864
$\mathbb{1}(age = 54)$	0.5368	0.0213	25.2417	$\mathbb{1}(age = 54)C_{54,t}$	1.5303	0.0722	21.1909
$\mathbb{1}(age = 55)$	0.5197	0.0215	24.1287	$\mathbb{1}(age = 55)C_{55,t}$	1.4709	0.0684	21.4962
$\mathbb{1}(age = 56)$	0.5311	0.0209	25.4126	$\mathbb{1}(age = 56)C_{56,t}$	1.4729	0.0678	21.7398
$\mathbb{1}(age = 57)$	0.5368	0.0204	26.3021	$\mathbb{1}(age = 57)C_{57,t}$	1.3766	0.0629	21.8955
$\mathbb{1}(age = 58)$	0.5520	0.0197	28.0636	$\mathbb{1}(age = 58)C_{58,t}$	1.4458	0.0652	22.1860
$\mathbb{1}(age = 59)$	0.5360	0.0197	27.2299	$\mathbb{1}(age = 59)C_{59,t}$	1.5187	0.0666	22.8119
$\mathbb{1}(age = 60)$	0.5176	0.0196	26.4157	$\mathbb{1}(age = 60)C_{60,t}$	1.4361	0.0609	23.5771
$\mathbb{1}(age = 61)$	0.5135	0.0191	26.9192	$\mathbb{1}(age = 61)C_{61,t}$	1.4710	0.0602	24.4286
$\mathbb{1}(age = 62)$	0.4959	0.0188	26.3625	$\mathbb{1}(age = 62)C_{62,t}$	1.6881	0.0659	25.6344
$\mathbb{1}(age = 63)$	0.4460	0.0191	23.3916	$\mathbb{1}(age = 63)C_{63,t}$	1.7310	0.0645	26.8438
$\mathbb{1}(age = 64)$	0.4201	0.0188	22.3774	$\mathbb{1}(age = 64)C_{64,t}$	1.8850	0.0660	28.5596
$\mathbb{1}(age = 65)$	0.3733	0.0187	19.9365	$\mathbb{1}(age = 65)C_{65,t}$	2.1040	0.0691	30.4674
$\mathbb{1}(age = 66)$	0.3020	0.0190	15.8836	$\mathbb{1}(age = 66)C_{66,t}$	2.3263	0.0721	32.2769
$\mathbb{1}(age = 67)$	0.2591	0.0188	13.7786	$\mathbb{1}(age = 67)C_{67,t}$	2.5799	0.0752	34.3078
$\mathbb{1}(age = 68)$	0.1961	0.0189	10.3919	$\mathbb{1}(age = 68)C_{68,t}$	3.0615	0.0841	36.3867
$\mathbb{1}(age = 69)$	0.1008	0.0194	5.1994	$\mathbb{1}(age = 69)C_{69,t}$	3.4420	0.0892	38.5728
$\mathbb{1}(age = 70)$	0.0210	0.0196	1.0709	$\mathbb{1}(age = 70)C_{70,t}$	3.7144	0.0911	40.7635
$\mathbb{1}(age = 71)$	-0.0406	0.0196	-2.0665	$\mathbb{1}(age = 71)C_{71,t}$	3.8462	0.0899	42.7952
$\mathbb{1}(age = 72)$	-0.0782	0.0194	-4.0391	$\mathbb{1}(age = 72)C_{72,t}$	4.0717	0.0926	43.9644
$\mathbb{1}(age = 73)$	-0.1163	0.0193	-6.0411	$\mathbb{1}(age = 73)C_{73,t}$	4.6079	0.1003	45.9347
$\mathbb{1}(age = 74)$	-0.1921	0.0195	-9.8549	$\mathbb{1}(age = 74)C_{74,t}$	5.1939	0.1068	48.6342
$\mathbb{1}(age = 75)$	-0.2837	0.0198	-14.3322	$\mathbb{1}(age = 75)C_{75,t}$	5.5554	0.1141	48.6800

Table 7: Regression results for eqn 3. R^2=.993 Fstat=2.39e3. p-value=0.

Using the estimated coefficients, we feed into equation 3 the projected college proportions for each age group and life expectancy estimates until 2150. Life expectancy estimates are assumed to be constant after 2090, which is the last year of the Social Security Administration's projected life tables. Our model predicts the 65 and over to be the most productive age group by 2035. Medical advances will continually improve where age 65 will no longer be the tail end of a person's life. The model's predictions are plausible as healthy bodily state and acquired experiences allow this group to be the most productive. Figure 41 shows the projections until 2150.

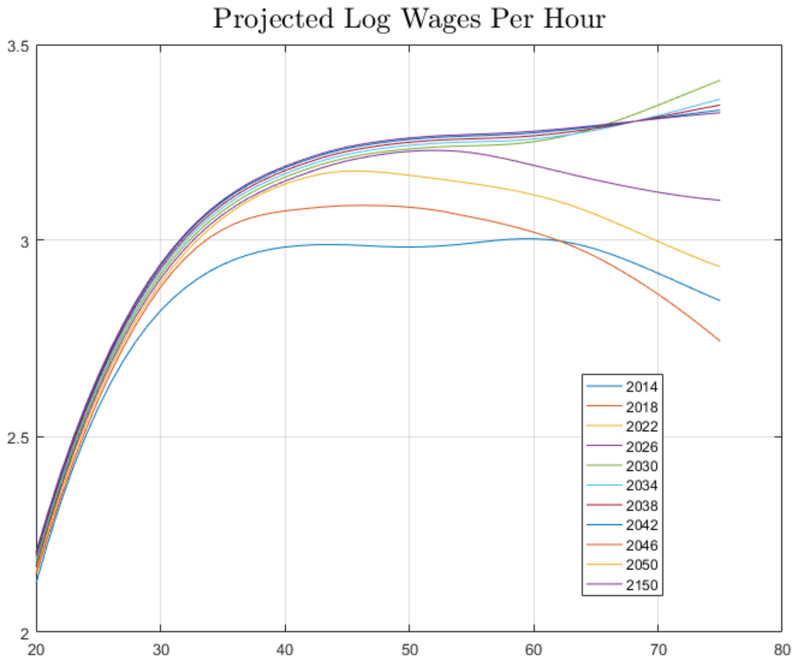

Figure 41: Data: ASEC CPS.

When the model is solved, each year of the transition will use its own projected log hourly wages as a proxy for the productivity profile. Population dynamics for each year is also needed. In theory, population evolution can be determined by survival rates and birth rates only. However, immigration, within year births and deaths make it difficult to match census data from birth and survival rates alone. Currently Census Bureau has population projections until year 2056. Social Security Administration has projections for births and survival rates until year 2090. Projecting the population with Social Security Administration's data under projects the population compared to the Census Bureau's projections. Figure 42 shows the difference in percentage terms between the two projections. Our method adjusts the population projections beyond 2056 by assuming the discrepancy $\phi_{t=2056} = \phi_{t>2056}$. For example, in year 2090, the number of 90-year-old are the number of 89-year-old in 2089 who have survived to age 90. But to account for the under estimation using only birth and survival rates, the population of 90-year-old will be adjusted upward 2%.

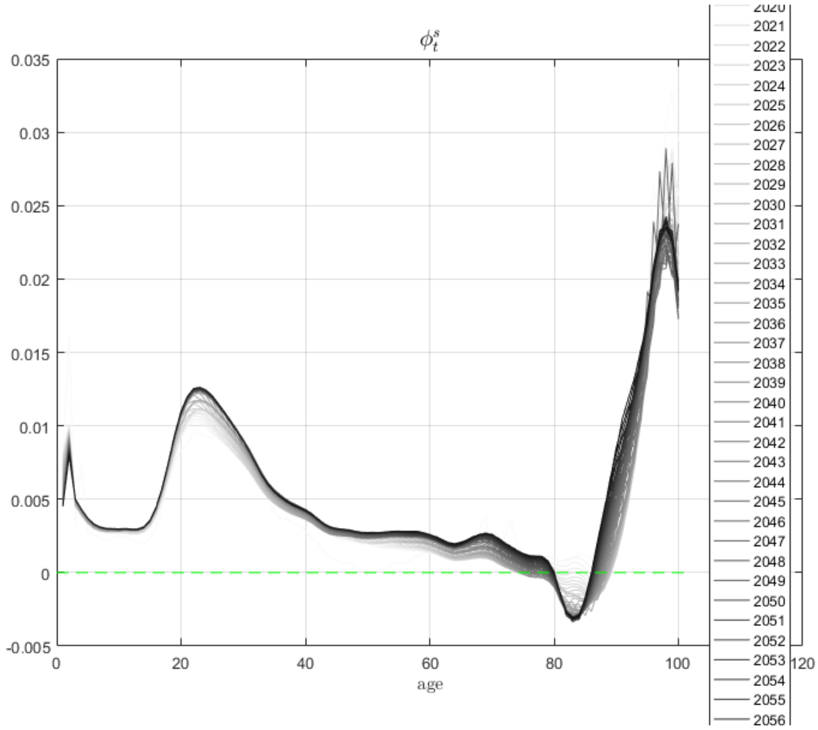

Figure 42: Data: US Census, SSA.

After 2090, the survival probabilities and birth rates do not change as we do not project these longer than the Social Security Administration's projections. Then the population starts toward a transition toward the steady state. Although these rates are constant the population will continuously evolve until finally reaching the steady state in 2150. Figure 43 shows transformation of the population density over the years. The density for each year will be used when aggregating the total labor and capital in the economy.

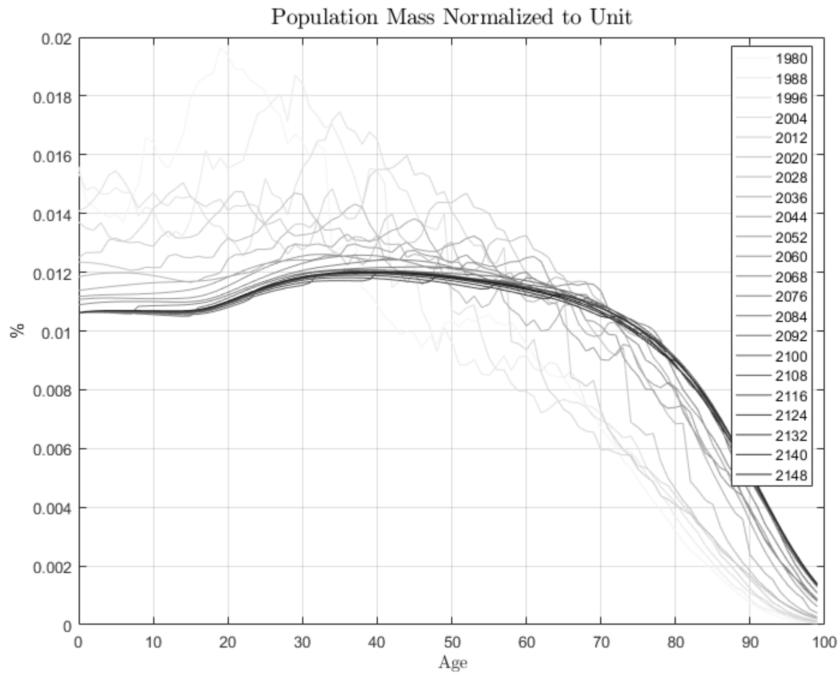

Figure 43: Data: US Census, SSA.

17 Model

The model economy is composed of consumers, perfectly competitive firms, and government. The economy at the stationary equilibrium is on a balanced growth path with constant birth rate, stationary survival probability and age-dependent productivity profile.

The economy is populated by overlapping generations of consumers. An agent enters the workforce at age 20, which we denote by $g = 0$. The maximum possible age is G (we assume $G = 79$), so that $g = 0, ..., G$. Let $\theta_{0,t}$ denote the measure of population that enters the workforce (i.e., is of age $g = 0$) at the beginning of date t, and let η_t be an exogenous, time-varying growth rate of this age group. Then, at date $t + 1$, the measure of age-0 generation is $\theta_{0,t+1} = (1 + \eta_t)\, \theta_{0,t}$. The agent of age $g < G$, who is alive at time $t + g$, survives to age $g + 1$ with probability $\psi_{g,t} > 0$; for $g = G$, we have $\psi_{G,t} = 0$. The probability of surviving to age $t + g$ at age g is $\pi_{g,t}^{g+t,t+g} \equiv \prod_{j=g}^{g+t-1} \psi_{j,t-g+j}$ where t is the date of entering the workforce; for $g = 0$, we use $\pi_{0,t} = 1$ for all t.

Consumers. An agent, who enters the workforce at date t, solves the following (detrended) utility-maximization problem

$$\max_{\{c_{g,t+g},n_{g,t+g},k_{g+1,t+g+1}\}_{g=0,\ldots,G}} \sum_{g=0}^{G} \widetilde{\beta}^{g}\pi_{g,t+g}\left[u\left(c_{g,t+g}\right)+\widehat{\varphi}_{g,t+g}v\left(1-n_{g,t+g}\right)+\left(1-\widehat{\varphi}_{g,t+g}\right)v\left(1\right)\right] \quad (79)$$

subject to

$$c_{g,t+g}+\gamma k_{g+1,t+g+1} = (1+r_t)k_{g,t+g}+(1-\tau_t)\widehat{\varphi}_{g,t+g}n_{g,t+g}e_{g,t+g}w_{t+g}+s_{t+g}\cdot \mathbf{1}_{\{g\geq G_R\}}+b_{t+g}, \quad (80)$$

$$n_{g,t+g} \geq 4 \text{ hours generates endogenous retirement decision.} \quad (81)$$

where $k_{0,t}=0$, and $k_{G+1,t+G+1}\geq 0$.

- $\widehat{\varphi}_{g,t+g}\leq 1$ is level of employment for agent of age g at time $t+g$. Some people do not want to participate in the market because they do not have good health or for some other reason. We assume that $\widehat{\varphi}_{g,t+g}$ changes over time because the health becomes and education becomes better and we will calibrate the change of $\widehat{\varphi}_{g,t+g}$ from the data - effectively, this is an exogeneous participation rate and we will project it .

Here, $c_{g,t}$, $k_{g+1,t+1}$ and $n_{g,t}$ are stationary consumption, end-of-period capital and hours worked of a g-year old at time t, respectively; the total time endowment is normalized to unity, so that $1-n_{g,t}$ represents leisure; r_t and w_t are the interest rate and wage per effective labor unit, respectively; γ is labor augmenting technological progress; τ_t is a Social Security tax; $\widetilde{\beta}$ is the detrended discount factor which will be specified later; $\varphi_{g,t}$ is the probability of being employed; $n_{g,t}$ is hours worked; $e_{g,t}$ is an age- and time-specific, exogenously given parameter; s_t is Social Security, received starting from the retirement age G_R (currently, it is $g=44$); $\mathbf{1}_{\{g\geq G_R\}}$ is an indicator function; b_t is an accidental bequest from the premature death.

We model age-dependent efficiency levels $e_{g,t}$ following Hansen (1993). We assume that the agent's labor efficiency is zero for any age greater than a working active age *G_active* (we assume it to be equal to 54). The aggregate efficiency level changes over time $\sum_{g=1}^{G} e_{g,t}=e_t$.

Addilog utility function. We consider the addilog type of utility function,

$$u\left(c,1-n\right)=\left[\frac{c^{1-\sigma}-1}{1-\sigma}+\phi\frac{\left(1-n\right)^{1-\eta}-1}{1-\eta}\right], \tag{82}$$

where σ, η, $\phi>0$; σ is the coefficient of risk aversion; η is the inverse of Frisch elasticity of labor. The system of FOCs for (82) is

$$\begin{aligned}\gamma c_{g,t+g}^{-\sigma} &= \widetilde{\beta}\psi_{g+1,t+g+1}\left[c_{g+1,t+g+1}^{-\sigma}(1+r_{t+g+1})\right],\\ \phi\left(1-n_{g,t+g}\right)^{-\eta} &= c_{g,t+g}^{-\sigma}(1-\tau_t)e_{g,t+g}w_{t+g},\end{aligned}$$

where $\widetilde{\beta}\equiv\beta\gamma^{1-\sigma}$.

Other version of the model with exogeneous participation rate This is anothe version of the model that can be checked. The difference is just in the FOC for labor.

An agent, who enters the workforce at date t, solves the following (detrended) utility-maximization problem

$$\max_{\{c_{g,t+g},n_{g,t+g},k_{g+1,t+g+1}\}_{g=0,\ldots,G}}\sum_{g=0}^{G}\widetilde{\beta}^{g}\pi_{g,t+g}\left[u\left(c_{g,t+g}\right)+v\left(1-\widehat{\varphi}_{g,t+g}n_{g,t+g}\right)\right] \tag{83}$$

subject to

$$\begin{aligned}c_{g,t+g}+\gamma k_{g+1,t+g+1} &= (1+r_t)k_{g,t+g}+(1-\tau_t)\widehat{\varphi}_{g,t+g}n_{g,t+g}e_{g,t+g}w_{t+g}+s_{t+g}\cdot\mathbf{1}_{\{g\geq G_R\}}+b_{t+g}, &(84)\\ n_{g,t+g} &\geq 4 \text{ hours generates endogenous retirement decision.} &(85)\end{aligned}$$

where $k_{0,t}=0$, and $k_{G+1,t+G+1}\geq 0$.

$$\begin{aligned}\gamma c_{g,t+g}^{-\sigma} &= \widetilde{\beta}\psi_{g+1,t+g+1}\left[c_{g+1,t+g+1}^{-\sigma}(1+r_{t+g+1})\right],\\ \phi\left(1-\widehat{\varphi}_{g,t+g}n_{g,t+g}\right)^{-\eta} &= c_{g,t+g}^{-\sigma}(1-\tau_t)e_{g,t+g}w_{t+g},\end{aligned}$$

Firm. At date t, a perfectly competitive firm solves

$$\max_{k_t, n_t} \quad k_t^{\alpha} n_t^{1-\alpha} - w_t n_t - (r_t - \delta)\, k_t, \tag{86}$$

where k_t and n_t are capital and labor inputs; $\alpha \in (0,1)$ is the share of capital in output; $\delta \in [0,1]$ is the depreciation rate.

The equilibrium factor prices are

$$r_t = \alpha \left(\frac{n_t}{k_t}\right)^{1-\alpha} - \delta, \tag{87}$$

$$w_t = (1-\alpha)\left(\frac{k_t}{n_t}\right)^{\alpha}. \tag{88}$$

Market clearing. Markets clear when aggregate capital and labor are the sums of all generations' capital and labor

$$k_t = \sum_{g=0}^{G} \theta_{g,t} k_{g,t}, \tag{89}$$

$$n_t = \sum_{g=0}^{G\_active} \overline{\varphi}_{g,t} e_{g,t} \theta_{g,t}, \tag{90}$$

where $\theta_{g,t}$ is the proportion of the population of age g at date t; we normalized it to $\sum_{g=0}^{G} \theta_{g,t} = 1$, and the aggregate efficiency skills at period t are $e_t = \sum_{g=0}^{G\_active} e_{g,t}$.

Government. The government's budget constraint is

$$\tau_t n_t w_t + \gamma \eta_t d_{t+1} = s_t \sum_{g=G_R}^{G} \theta_{g,t} + (1 + r_t^d) d_t. \tag{91}$$

Note the government's budget does not need to be balanced in every period. The balanced budget is where the Social Security payments adjust accordingly to tax receipts. In this setup, the government can promise fixed social-security payments for a period. The government debt d_t evolves with a lower interest than the return on capital $r_t > r_t^d$ in order to ensure that the debt does not grow too fast. The interest rate, charged on the government's debt, is chosen exogenously.

Accidental bequests. The accidental bequest payout is

$$\gamma\eta_t b_t \sum_{g=0}^{G-1} \theta_t^g = \sum_{g=1}^{G-1} (1-\psi_{g-1,t-1})\theta_{g-1,t-1}k_{g-1,t-1}. \tag{92}$$

Budget constraint in life-time form.

18 Results

The model is simulated two ways: 1) using static profiles of wage-age and employment percentage-age 2) using dynamic profiles of wage-age and employment percentage-age. When the model is solved using static profiles we set the wage-age and employment participation age profiles to that of 2014 for all years. In order to compare the evolution of the Social Security Trust Fund between the two, we set the pension to be the same in both simulated paths. The utility function is parametrized as log for consumption and leisure. The parameter values below (table 8) are common in the literature. The dis-utility of labor was set such that labor choice during peak working age was around .35.

Parameter	Description	Value
σ	coefficient of risk aversion	1
η	inverse frisch elasticity	1
ϕ	dis-utility of labor	2.5
α	capital share	.36
β	discount factor	.96 (annual)
γ	growth rate	.02
τ	social security tax	.14
δ	depreciation	.07

Table 8: Calibration Values

The evolution of the Social Security Trust Fund is shown in Figure 44 for both the static and dynamic paths. Both simulated paths show the fund will drastically experience sharp decreases in annual balances. The bottom left of Figure 44 shows the annual balance of Social Security Fund: receipts (top) minus payouts. The payouts are not included in the figure since both simulated paths use identical payouts. If the model is calibrated correctly, the bottom left of Figure 44 would reflect a deficit but would feature the same sharp decrease. Note the differences are negligible. In terms of stock of surplus or deficit, the difference between two simulated paths are quite insignificant.[20] Although the difference in the state of the Social Security Trust Fund is negligible, the economic states between the two simulated paths are worth noting.

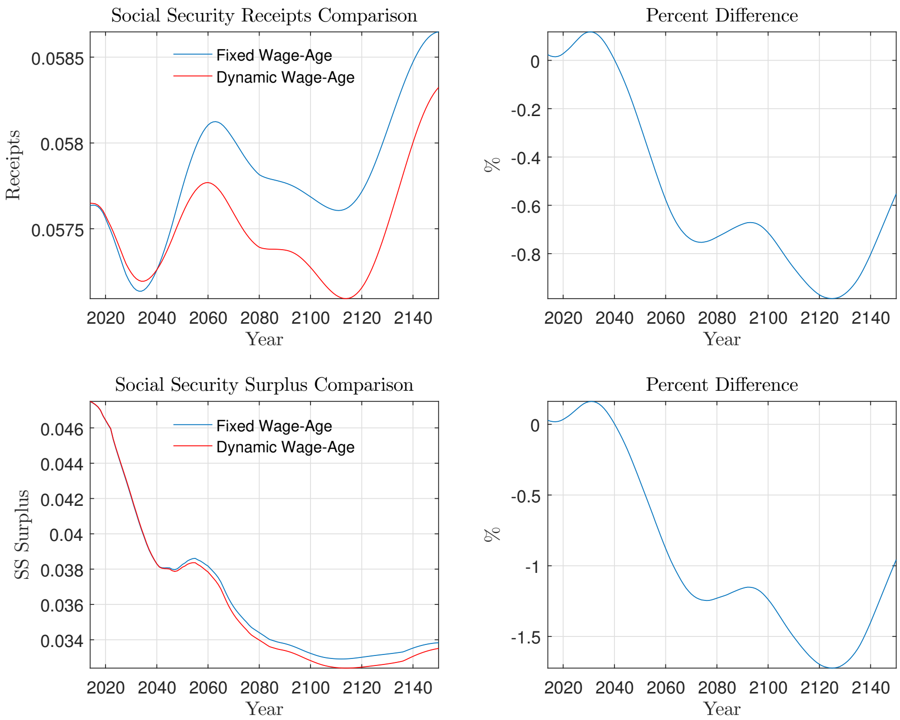

Figure 44: Social Security Fund States.

In Figure 45 the aggregate states of both simulated paths are shown. Top row shows the efficiency weighted aggregate labor in the economy. Dynamic profile path has higher total labor in the economy. This is expected since the dynamic profiles incorporate increasing productivity as the population become more educated. Higher total labor in the economy should imply more receipts for the Social Security Trust Fund. However, wages (Figure 46) are declining in the dynamic profile path as the aggregate capital stock (bottom 45) in the economy is lower. The changing productivity profile also affects savings decisions for the agents.

[20] At year 2035, the difference between the two stocks will be less than 25% of one year's expenditures.

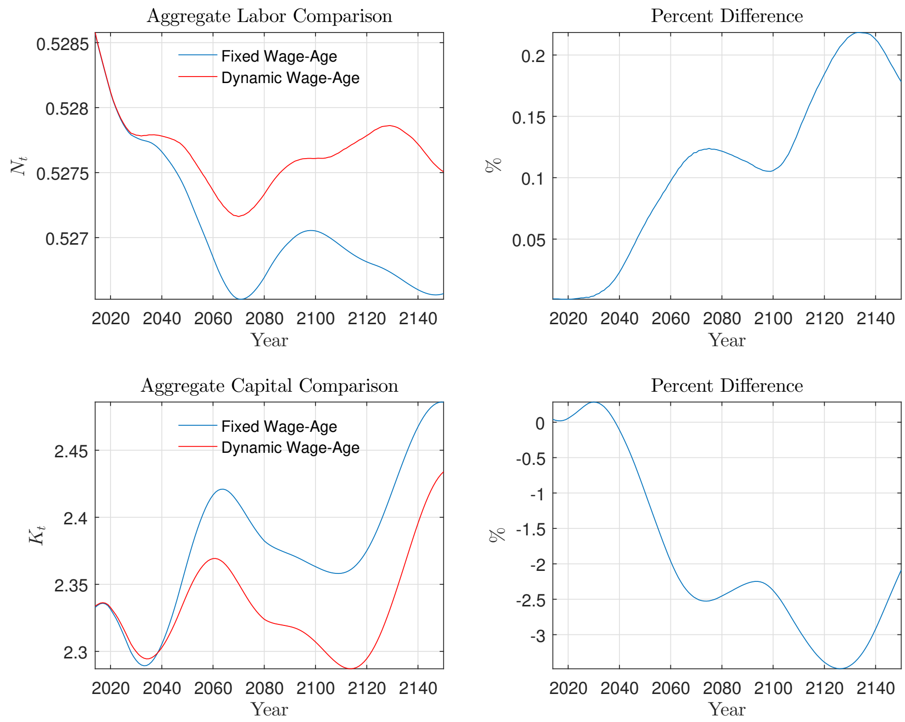

Figure 45: Aggregate States

In middle row of Figure 47 the capital choices at the final steady state for both simulated paths are shown. Note in the dynamic profile path, the younger agents have less capital but older agents have more capital than the agents in the static profile path. In the dynamic path, younger agents face declining productivity therefore they have less income for savings. These agents are more productive in later years. Agents are choosing to work more during the productive years as shown in top row Figure 47, essentially substituting labor between time. But the population density is not uniform, that is there are significantly more young than the old. This lowers the aggregate capital stock in comparison to the static profile transition. The decrease in capital stock essentially offsets the increases in total labor of the dynamic profile path, resulting in negligible changes to the state of Social Security Trust Fund.

The bottom row of Figure 47 shows the consumption for the agents at the final steady state for both simulated paths. Note the consumption is higher for the old using the dynamic profile. Some of state variable, individual or aggregate are sensitive to calibrated parameters. However, aggregate capital and individual capital differences and consumption profile seem to hold for most calibration parameters.

The dynamic profile transition strengthens the case for reducing benefits to reform Social Security Trust. The transition paths imply the OLG models using static profiles will underestimate the wealth of older agents in future years.

Decreasing benefits will be highly unpopular. Another unpopular alternative is to increase Social Security tax or the tax base as in inclusion of capital income. The simulated paths show fluctuations in aggregate capital and labor. Proposed tax increases should take these fluctuations into account as to minimize the impact on aggregate labor and capital.

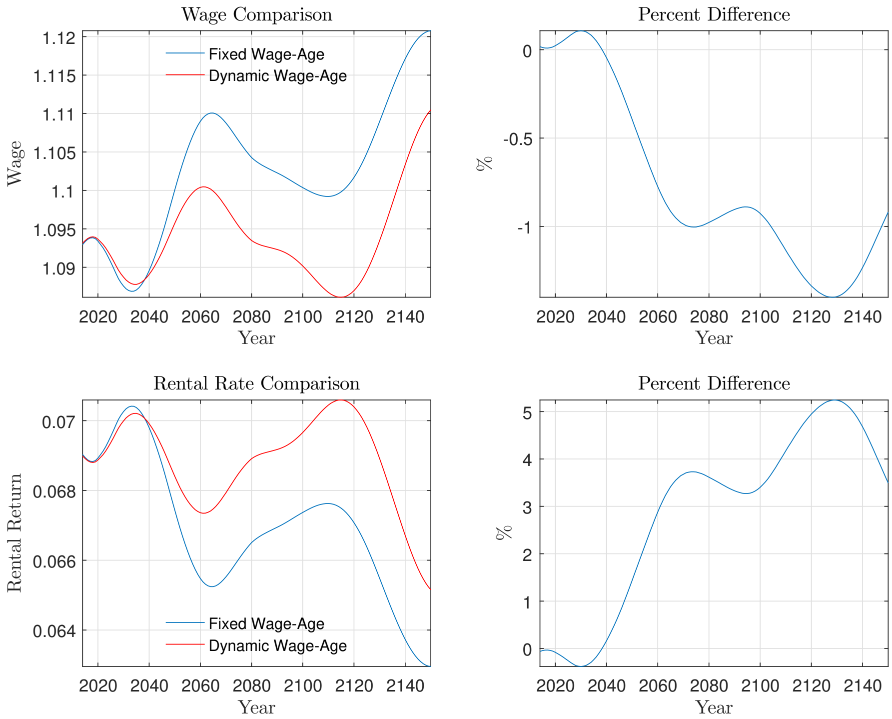

Figure 46: Prices

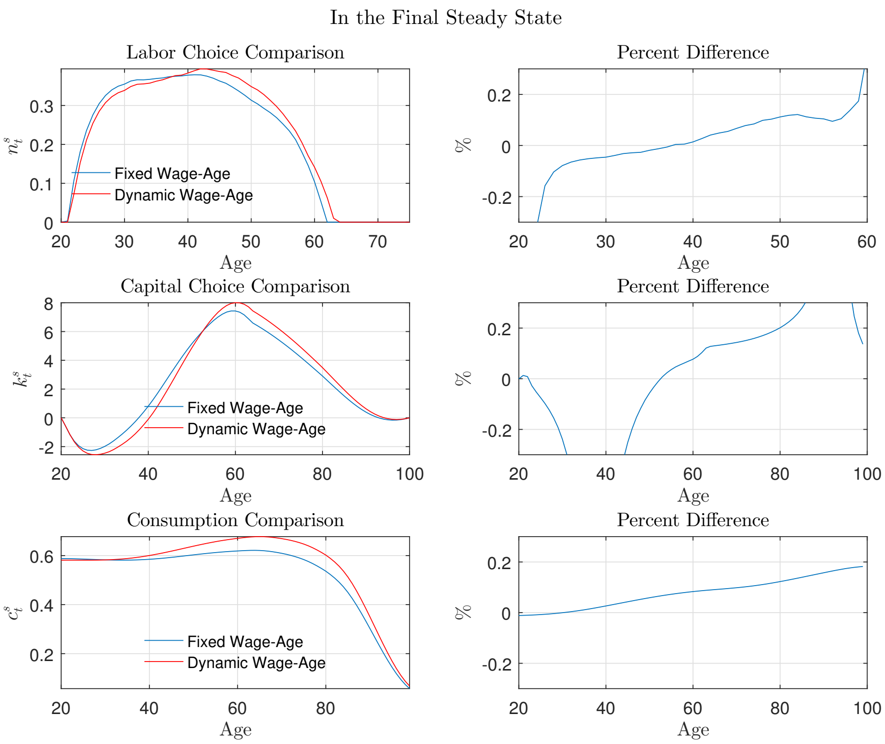

Figure 47: Prices

19 Conclusion

The paper extrapolates the current labor patterns using education as a proxy. We solve the transition dynamics using dynamic wage-age and employment participation-age profiles. The difference in the state of the Social Security Trust Fund is negligible compared to the static profile transition. But the differences in economic states such as individual and aggregate capital are noteworthy. These differences should be taken into account when or if policy is constructed to reform the Social Security Trust Fund.

ïz£

References

Auerbach, A. and Kotlikoff, L. (1987), "Dynamic Fiscal Policy," .

Chetty, R., Guren, A., Manoli, D. S., and Weber, A. (2011), "Does Indivisible Labor

Explain the Difference Between Micro and Macro Elasticities? a Meta-Analysis of Extensive Margin Elasticities," *NBER Working Paper*.

de la Croix, David, Pierrand, Olivier, Sneessens, H. (2010), "Aging and Pensions in General Equilibrium: Labor Market Imperfections Matter," *Journal of Economic Dynamics and Control*.

Heckman, James. (1984), "Comments on the Ashenfelter and Kydland Papers," *Carnegie-Rochester Conference Series on Public Policy*.

Hansen, G. D. (1993), "The Cyclical and Secular Behaviour of the Labour Input: Comparing Efficiency Units and Hours Worked," *Journal of Applied Econometrics*, 8, 71–80.

Lemieux, T. (2006), "The Mincer equation thirty years after schooling, experience and earnings," *Jacob Mincer. A Pioneer of Modern Labor Economics*, Part IV pp 127-145.

McGrattan, E., Prescott, E. (2015), "On Financing Retirement with an Aging Population," *Federal Reserve Bank of Minneapolis Staff Report 472*, 53, 160.

National Research Council (2012), "Aging and the Macroeconomy. Long-Term Implications of an Older Population." .

Nishiyama, S. (2013), "Fiscal Policy Effects in a Heterogeneous-Agent Overlapping-Generations Economy With an Aging Population," *Journal of Economic Dynamics and Control*, 61, 114–132.