

Sensitivity

An inquiry into the robustness of model calibration as used in modern
macroeconomic research

John Sæten Lilletvedt

Master's Thesis

The thesis completes a

Master's degree in Economics

University of Bergen, Department of Economics

[June 2021]



UNIVERSITETET I BERGEN

Acknowledgments

I would like to thank my supervisor Eirik Strømland for suggesting the broader theme of my research question, for invaluable guidance and encouragement, and for pushing my analyses and my writing to much higher levels than what I had hoped for when I first began work on the thesis. My parents, my sister, and my closest friends have all helped me by reading the thesis, asking poignant questions about the content, and correcting my writing. I am incredibly thankful for their support. Any errors, linguistic or otherwise, are entirely my own.

Abstract

Calibration is a much used but problematic method for achieving quantitative predictions from modern macroeconomic models. As with other quantitative work in economics, there is reason to believe that findings on the basis of calibration can be subject to measurement error, specification error, and biases. To investigate the robustness, replicability, and credibility of findings based in calibration, I replicate and extend a model of the recent Chinese growth experience. The extensions are based on the principles of sensitivity analysis as suggested by Edward Leamer (1983) with the explicit goal of illustrating the possible uncertainty about the model results. In the case of the particular model investigated here, the results are reasonably robust to alternative calibrations as long as one excludes any alternative calibrations that violate the theoretical assumptions. There are calibrated parameters for which the model is much more sensitive to alternative specifications.

Table of contents

1. INTRODUCTION	1
2. CALIBRATION – ORIGINS AND CRITIQUE.....	4
2.1 THE CASE FOR CALIBRATION	4
2.2 PROBLEMS WITH CALIBRATION.....	6
3. A PRESENTATION OF THE REPLICATED MODEL.....	9
3.1 THE MODEL	11
3.1.1 Preferences and technology	11
3.1.2 Equilibrium during transition.....	13
3.1.3 Foreign surplus, savings and investment	14
3.1.4 Equilibrium post-transition and financial development	15
4. REPLICATION, METHODOLOGY, DATA AND CONDITIONS FOR THE ANALYSIS	15
4.1 THE ORIGINAL CALIBRATION	16
4.2 METHODOLOGY	20
4.3 ESTIMATES USED FOR EXTENSIONS	21
4.3.1 The capital share of income in China	21
4.3.2 The capital depreciation rate in China	25
4.3.3 The conundrum of the estimated intertemporal elasticity of substitution	27
4.3.4 The estimated elasticity using data from China and “best practice” estimates	29
4.4 MODEL ASSUMPTIONS AND VIOLATIONS	30
5. ANALYSIS	33
5.1 EXTENSIONS AND ROBUSTNESS CHECKS	34
5.1.1 The model subject to all possible calibrations given alternative estimates	34
5.1.2 The model subject to assumption-consistent alternative calibrations	36
5.1.3 Model robustness subject to incremental changes of the benchmark values	38
6 CONCLUSION	44
LITERATURE	44
APPENDIX.....	52
A - GROWING LIKE CHINA PRESENTED ANALYTICALLY	52
B – EXAMPLE OF CODE FOR INCREMENTAL TEST	61

1. Introduction

Quantitative (social) science has in recent years gone through a replication crisis: Attempts to replicate studies in for instance psychology have shown that only a small share of significant effects are replicable (Open Science Collaboration, 2015). Replication has also been a point of discussion in economics, and for quite a while (Dewald et al., 1986). A study of the *Journal of Money, Credit and Banking* archive found that only 22% of replications gave the same results as the original articles (McCullough et al., 2006). A more recent effort to replicate studies from experimental economics find a more reassuring replicability percentage of 61% on average (Camerer et al., 2016). Questions of replicability beg questions of scientific credibility, meaning questions about the manipulation of findings. Credibility has been a theme in economics and especially econometrics for even longer than replicability (Angrist & Pischke, 2009, 2010; Glaeser, 2006; Leamer, 1983, 2010). Joshua Angrist and Jorn-Steffen Pischke (2010) claim that applied econometrics has seen no less than a credibility revolution since the 1980s and so believe the issue to be in part solved, at least in some fields of econometric research. Edward Leamer (2010), on the other hand, maintains that his original critique of econometrics still holds: A finding may hinge upon some choice done by the researcher (Leamer, 1983, pp. 36–38).

One field of economics where Angrist and Pischke agree with Leamer that credibility has not been addressed in a satisfactory way, is macroeconomics (Angrist & Pischke, 2010, pp. 18–20; Leamer, 2010, p. 44). In macroeconomics, much quantitative work is being done within the dynamic stochastic general equilibrium paradigm (Finansdepartementet, 2019; He et al., 2017; Jones & Klenow, 2016; Smets & Wouters, 2007; Song et al., 2011). Models are built on the basis of general equilibrium theory and founded in microeconomic behavior to produce predictions about the effects of macroeconomic policy (Kydland & Prescott, 1982; Lucas, 1976). The models are often quite complex. Given model complexity, researchers usually do what is known as a calibration to get a clear numerical prediction from their model. Calibration is the practice of specifying numerical values for some or all parameters in a model, based mostly on empirical work in other fields of economics (Hansen & Heckman, 1996; Kydland & Prescott, 1982, 1991, 1996). The need for calibration and the sheer complexity of the models have made the general equilibrium research program subject to a lot of criticism (Angrist & Pischke, 2010; Blanchard, 2016; Hansen & Heckman, 1996; Leamer, 2010; Romer, 2016). Paul Romer (2016) holds that: “For more than three decades, macroeconomics has gone backwards.

The treatment of identification is no more credible than in the early 1970s but escapes challenge because it is so much more opaque.” The assumptions underpinning the standard macroeconomic model today may be as unrealistic and incredible as in earlier research but are now hidden under layers of mathematical functions and technical jargon which make them hard to spot and critique (Romer, 2016, p. 12). The practice of calibration is one part of the opacity of the models as it is used to justify specific numerical values for solution parameters in a somewhat arbitrary (Canova & Sala, 2009, p. 435) and non-transparent way.

Motivated by the concerns about transparency in the general equilibrium paradigm in macroeconomics, the present thesis is an inquiry into a particular case of calibration. I replicate a much cited article which uses calibration and is published in a journal with a high article influence score – *Growing Like China* by Zheng Song, Kjetil Storesletten and Fabrizio Zilibotti (2011) – and subject the model to a number of tests by specifying alternative values for the calibrated parameters. The extensions I do are based on the idea that the published estimates used for calibration may undermine the uncertainty about model findings. By using relevant alternative estimates we get a picture of the degree of uncertainty associated with the model predictions (Leamer, 1983; Simonsohn et al., 2020). The research question I ask is: Are the quantitative results in Song et al. (2011) robust to alternative calibrations? The alternative calibrations are based on other research that is relevant to the question Song et al.’s model is supposed to answer. Such tests are in line with the method of calibration as originally proposed (Kydland & Prescott, 1982, 1991, 1996) and as it is practiced today (Canova & Sala, 2009; DeJong, 2011).

My goal is to bring the broader debate about replication and credibility to the specific form of macroeconomic research that is dynamic stochastic general equilibrium models. In part I aim to investigate the strong claims of those who criticize the models (Blanchard, 2016; Romer, 2016) and in part I aim to cover what I consider a blind spot in the current macroeconomic research field. The conceptual problems of calibration have been addressed before (Beltran & Draper, 2008; Canova & Sala, 2009), but there are to my knowledge no studies that replicate and extend model findings based in calibration to check for robustness, manipulation, and measurement error. The thesis aims to initiate a broader investigation into the robustness and replicability of findings based in calibration.

The model I replicate and extend attempts to explain some distinct features of the Chinese growth experience that are puzzling in the light of neoclassical growth theory: High capital returns and a growing foreign surplus along with high economic growth (Song et al., 2011). High capital returns and a growing foreign surplus has been observed for many countries and been named an allocation puzzle (Gourinchas & Jeanne, 2013). Song et al. attempt an explanation of the Chinese growth experience through the general theory of input misallocation: There may be frictions that keeps for instance workers in low-productivity firms and if frictions disappear one will see higher productivity growth due to reallocation (Diao et al., 2017). Before going into detail about the way Song et al. use misallocation to explain Chinese growth, I present the cases for and against calibration.

2. Calibration – Origins and Critique

2.1 The case for calibration

Calibration as practiced today dates back to the original effort to build macroeconomic models founded in microeconomic behavior in line with Robert Lucas's critique of macroeconomic policy evaluation (1976). Edward Prescott and Finn Kydland presented their real business cycle theory as a step towards microfounded models in *Time to Build and Aggregate Fluctuations* (1982). To evaluate the new type of model quantitatively, Kydland and Prescott suggested calibration (1982, pp. 1360–1363). Kydland and Prescott are not especially clear on what calibration entails in *Time to Build...* How calibration differs from estimation is also left unaddressed (Hansen & Heckman, 1996, p. 92). Prescott made the calibration methodology more explicit in a later conference paper:

The beauty of this model is that both growth and micro-observations can be used to determine its production and utility functions. [...] The key parameters of the model are the intertemporal and intratemporal elasticities of substitution. As Lucas (1980, p. 712) emphasizes, “On these parameters, we have a wealth of inexpensively available data from census cohort information, from panel data describing the reactions of individual households to a variety of changing market conditions, and so forth.” To this list we add the secular growth observations which have the advantage of being experiments run by nature with large changes in relative prices and quantities and with idiosyncratic factors averaged out. (1986, p. 22)

In other words, estimates of parameter values are taken from relevant research and plugged into the model to achieve a model solution.

Calibration is proposed by Kydland and Prescott as being distinct from estimation. Rather than being part of standard probabilistic econometrics, calibration is held to fall within a broader sense of the econometric field in line with Ragnar Frisch's original motivation for founding the journal *Econometrica*: The unification of statistics, mathematics, and theory to answer economic questions (1933b, p. 2). Kydland and Prescott argue that their approach to macroeconomics should be viewed as consistent with econometrics as it was first envisioned

by Frisch, Irving Fisher and Joseph Schumpeter (Kydland & Prescott, 1991). How can the general equilibrium approach of Kydland and Prescott represent a unification as envisioned by Frisch when they explicitly dismiss statistical estimation as relevant? The dismissal is in fact practically motivated. Kydland and Prescott argue that the level of abstraction of their models necessarily make them false, and the model predictions will be dismissed by adherence to standard notions of significance and fit (1991, 1996). Abstraction does not make the models unusable as long as the models are motivated by clear-cut research questions¹. The motivation for proposing what Kydland and Prescott call *computational experiments*² as an alternative econometric tool is put forward as a case for pluralism in the quantitative treatment of economic theory. Calibration is a necessary part of such computational experiments.

Why is calibration necessary? Kydland and Prescott illustrate the idea of calibration by likening economic models to thermometers (1996). For a thermometer to measure the temperature accurately, one must establish some facts about temperature, such as when water freezes. In economics there are similarly some facts about the workings of a particular economy or the like which a model must account for if it is to give a convincing answer to a research question: ‘Thus, data are used to calibrate the model economy so that it mimics the world as closely as possible along a limited, but clearly specified, number of dimensions.’(1996, p. 74). Calibration is used as a way to achieve credible answers to the clear-cut research question at the center of the model. In less metaphorical but similar terms, one can rationalize any process as equilibrium behavior for a model built on the principles of Kydland, Prescott, and Lucas unless one restricts the model in some way (Canova & Sala, 2009, p. 433; Kydland & Prescott, 1991, p. 169). Calibration is used to ensure model behavior consistent with theory.

Calibration can also help with some other problems that may occur when finding a solution to the general equilibrium macro models. Most such models consist of a system of non-linear equations and should be solved as a system. To get one specific solution one needs to approximate. When approximating, some parameters that feature in the general model can

¹ “Can a neo-classical growth model account for historical business cycle fluctuations in the US if the only source of shocks is technological transformation?” (Kydland & Prescott, 1982, 1991, 1996)

² The computational experiment, Kydland and Prescott hold, can in fact be viewed as a form of estimation (1996, p. 74). For the computational experiment to be possible, the model must have a solution and thus must be calibrated. It is not exactly clear how calibration differs from estimation when it is a necessary part of the computational experiment.

disappear from the solution (Canova & Sala, 2009, p. 433). In some cases, a parameter may not disappear from the solution all together but rather only appear in the solution proportionally to another parameter (Canova & Sala, 2009, p. 433). Additionally, for the model results to be robust the solution must in some way be dependent on the parameters of the model. Dependence will show up as curvature of the plotted model solution across the parameter space. One then has a problem if there is little or no curvature of the solution for certain parameters (Canova & Sala, 2009, p. 433). Calibration is a possible solution to the problems because the problematic parameters are pinned down by specific numerical values (Canova & Sala, 2009, pp. 433–435).

2.2 Problems with calibration

Calibration has been quite widely criticized, both implicitly as part of a larger critique of the general equilibrium research program (Blanchard, 2016; Canova & Sala, 2009; Romer, 2016) and more specifically as a method (Andersen, 1991; Hansen & Heckman, 1996). Considering the issues one may encounter when approximating a model solution, calibration in combination with mutually dependent parameters can be generally problematic (Canova & Sala, 2009, p. 435). If the estimates used for calibrating the model are in some way incorrect or biased, the calibration will bias the numerical results when some parameters only feature in the model solution as pairs or proportionally to each other.

We have reason to be concerned about sources of error in the estimates used for calibration. In practice some parameters are calibrated through the researcher's educated guess about estimates (Canova & Sala, 2009, p. 435), a method often used in earlier macroeconomic research programs³ (Frisch, 1933a; Hansen & Heckman, 1996; Romer, 2016, p. 12), rather than calibration as exclusively based on relevant research (Prescott, 1986, p. 22). Educated guesses can of course be wrong and empirical estimates are subject to measurement error. Economic

³ The proponents of calibration argue on the optimistic side that we do have enough data for models to give correct predictions from estimates alone (Kydland & Prescott, 1991, 1996; Lucas, 1980; Prescott, 1986). The actual practice of research may retain some practices from early simulation such as in Frisch (1933a), where calibration was done simply by guessing. Guessing is however counter to the initial motivation for reviving Frisch's method (Kydland & Prescott, 1991, p. 163). An interesting line of investigation, which in part will be followed here, is to investigate whether educated guesses are justified due to a lack of data or whether there are other considerations at play when estimates are guessed at.

models are not like thermometers as long as some economic facts are more controversial than what is assumed in calibration exercises (Blanchard, 2016).

The practice of researchers sometimes guessing at the estimates used for calibration makes for a link between the controversy of calibration and the broader debate about the replication and credibility crisis in empirical science. There is overwhelming evidence that researchers, knowingly and unknowingly, often manipulate empirical findings to fit their claims (Brodeur et al., 2016; Simmons et al., 2011; Simonsohn et al., 2020). For instance, the disclosure standards that have been practiced until recently can be shown to make false positive findings much more likely (Simmons et al., 2011). Journals are less likely to publish null findings or replications, which biases research towards a focus on significance while there is too little focus on power (Brodeur et al., 2016; Duvendack et al., 2017; Franco et al., 2014; Simmons et al., 2011). One possible explanation of the problem may be confirmation bias (Simmons et al., 2011), which is also a likely cause of bias in journals and among referees (Glaeser, 2006; Havránek, 2015; Ioannidis, 2005).

The replication crisis, as well as being prevalent in quantitative studies in general, applies to fields where the data is generated from experiments, i.e., research where there should be no lack of relevant data (Simmons et al., 2011). It is reasonable, then, to assume that the problems of replication and credibility in empirical science apply to the general equilibrium approach in macroeconomics as well, especially considering the practice of researchers guessing at estimates used for calibration (Canova & Sala, 2009, p. 435). Biases will be present too when estimates are taken from data exclusively, but one would think the biases amplified by the researchers having to guess when data is not available. If the computational experiment and calibration is to be understood as econometrics (Kydland & Prescott, 1991, 1996), Leamer's critique of econometrics applies: Reported results may hinge upon some choice done by the researcher (1983, pp. 36–38).

Calibration as a method also raises concerns about measurement and specification error. Lars Peter Hansen and James Heckman say it best: “It is simply not true that there is a large shelf of micro estimates already constructed for different economic environments that can be plugged without modification into a new macro model.” (1996, p. 100).

First and foremost, there is a mismatch between the environments invoked to justify microeconometric estimation and the environment of the general equilibrium models (Andersen, 1991, pp. 181–182; Hansen & Heckman, 1996, p. 94). Consider the work of Johns Shoven and Whalley on deterministic general equilibrium models (1992). Calibration as they present it is fraught with problems, such as the estimates of elasticities used for calibration being based on quite spares, differing, and often contradictory studies⁴ (1992, p. 105). Moreover, calibration is generally based on one observation for each parameter, often a time series average, which may mean that a unique set of values for the parameters of a model is not found from the estimates used. Hansen and Heckman believe that the concern about relevant and usable data for calibration is as strong, if not stronger, when the models are stochastic (Hansen & Heckman, 1996, p. 94).

The practice of using sample averages for calibration tends to go along with a lack of reportage of sample error or relevant measures of correlation (Hansen & Heckman, 1996, p. 95). Exclusion of the sample error of the averages or other relevant statistics can only be justified in special cases: A long-run average must be representative of the short-run dynamics of the model as well. Assumed short-run stability will only hold in very special circumstances (Hansen & Heckman, 1996, p. 95). Furthermore, calibration does not entail matching the full set of dynamics of the model to the full set of dynamics of the data (Hansen & Heckman, 1996, p. 96; Oreskes et al., 1994, p. 643). There is then a risk of verification being indistinguishable from the initial calibration, which makes the estimated results directly depend on the calibration. The distinction between calibration and estimation is not clear (Blanchard, 2016, p. 2) and the problem is likely only more prevalent for more sophisticated models (Glaeser, 2006, p. 18). Biased calibration will not disappear through combination with other techniques such as Bayesian estimation (Blanchard, 2016; Canova & Sala, 2009).

The critique of calibration implies two criteria for it to be a legitimate method for identification: The use of any numerical value for calibration must be argued for in detail, and there must be alternative specifications to check the robustness of the quantitative implications of the model (Canova & Sala, 2009; Glaeser, 2006; Leamer, 1983; Simmons et al., 2011; Simonsohn et al.,

⁴ Kydland and Prescott would contest the claim: ‘Because the language used in these business cycle models is the same as that used in other areas of applied economics, the values of common parameters should be identical across these areas and typically have been measured by researchers working in these other areas.’ (1991, p. 170). They then go on to argue that calibration is actually the forte of their approach.

2020). A replication must consider and evaluate the justifications for the benchmark calibration and any robustness checks in the original article. I add that in replicating one must note any assumptions about the analytic model, as such assumptions could imply upper or lower bounds for the values used for calibration. A detailed presentation of the model in *Growing Like China* (Song et al., 2011) is therefore in place.

3. A presentation of the replicated model

The model I replicate is an attempt to explain the puzzle of a high growth rate and high return on capital contra a positive and increasing foreign reserve surplus in the Chinese economy (Song et al., 2011). Neoclassical theory predicts falling returns to capital if the investment rate in an economy is high, but the return to capital has remained high in China despite an increasing investment rate (Song et al., 2011, pp. 196–197). Theory also predicts a net flow of capital towards an economy where the returns to capital are high, but China has run a foreign surplus at least since the 1990s (Song et al., 2011, p. 197). China is by no means the only developing country that runs a foreign surplus and enjoys high levels of growth: Pierre-Olivier Gourinchas and Olivier Jeanne (2013) have documented the break with neoclassical theory and named it the allocation puzzle. Song et al. seek to amend neoclassical theory to include the allocation puzzle (2011, p. 197). Their proposed amendment is to allow for systematic differences in firm level productivity and misallocation of inputs.

Song et al. assume an environment with two types of firms: Entrepreneurial, E, and financially integrated, F. E-firms are more efficient in their use of inputs, modeled as a productivity advantage. F-firms survive because of financial frictions. There are barriers for E-firms to receive funding from the financial sector and any E-firm investment is dependent upon internal savings. A sufficiently high saving rate among managers of E-firms will lead to a transition where labor reallocates towards E-firms. As F-firms disappear due to reallocation the demand for external financing declines and the high saving rate is invested abroad leading to an increasing foreign surplus.

E- and F-firms are meant to illustrate the difference between domestic private enterprises and state-owned enterprises in China. Song et al. argue that there are considerable financial frictions that affect private enterprises more than state-owned enterprises (2011, pp. 201–204). One main fact to underline the differing financial integration is that production in state-owned enterprises is capital-intensive while private enterprises have largely taken over in labor-intensive production.

3.1 The model

3.1.1 Preferences and technology

People live for two periods in an overlapping generations environment. In the first period, people work and in the second period they live off of their pension. The workforce equals the general population and so grows at the same rate. A fraction of each cohort are entrepreneurs. The entrepreneurs have higher skills than the workers and these skills are perfectly transmitted from one generation to the next. Due to entrepreneurial skills, E-firms have a productivity advantage. General productivity grows at an exogenous rate and capital fully depreciates for each period.

Inherited skills lead to more productive E-firms:

$$y_{Ft} = k_{Ft}^\alpha (A_t n_{Ft})^{1-\alpha}$$

$$y_{Et} = k_{Et}^\alpha (\chi A_t n_{Et})^{1-\alpha}$$

The productivity difference between E- and F-firms arises from a choice of how to manage production: Either delegate management or maintain direct control over all strategic decisions. The choice is associated with a trade-off. Delegation gives higher productivity in the form of $\chi > 1$, but there is the risk of theft of income by the manager. By paying the managers a higher share than he can possibly steal, theft may be averted. The inherited skill of an entrepreneur is to be better at monitoring her manager, making it impossible for the manager to steal more than a fraction of income, $\psi < 1$. The difference in productivity therefore arises from different choices of corporate governance: E-firms will always choose delegation, while F-firms will always choose management by owners.

There are two types of laborers: Workers and managers. Workers earn the competitive wage rate while managers earn a fraction of the income of the firm they manage. While workers only have the choice of saving some of their wage income in a bank account, managers can choose either to save their managerial income in a bank account or to save by reinvesting in the firm

they manage. What is saved in bank accounts is reinvested by the financial sector either in domestic firms or in foreign bonds. The output of E-firms is assumed to be non-verifiable, and entrepreneurs are only able to borrow the share that they can guarantee to repay in period two, making for a credit constraint. There are considerable financial frictions as well due to operational costs, red tape etc. amounting to an iceberg cost of lending to domestic firms. The iceberg cost makes for a higher interest rate on loans than on savings.

To determine how managers will allocate their savings, we must determine an expression for the returns of E-firms. Returns are found by standard profit maximization in F-firms which establishes the wage rate and in turn enables value maximization for E-firms. Reallocation of labor from F- to E-firms is only assured if the rate of return to E-firms is higher than the interest rate on loans. E-firm productivity must therefore surmount the loss of E-firm income due to managerial compensation.

ASSUMPTION 1:

$$\chi > \underline{\chi} \equiv \left(\frac{1}{1 - \psi} \right)^{\frac{1}{1-\alpha}}$$

Assumption 1 assures that E-firms prefer delegation over centralization as a model for corporate governance and that young entrepreneurs prefer to invest in E-firms. A violation of the assumption would mean no E-firms in equilibrium. Assumption 1 is only binding for a sufficiently severe credit constraint on E-firms. The saving rate of entrepreneurs is found by noting that E-firm capital is given by the sum of savings and the amount E-firms can borrow.

We will do well to note some of the assumptions made so far. All firms are, regardless of productivity, assumed to be competitive in the benchmark model and the labor market is competitive and frictionless. Neither assumption is realistic, but they enable focus on the frictions at the base of the model (Song et al., 2011, p. 211). As for the difference between the two types of firms, there is some evidence that delegation of management is uncommon in state-owned enterprises relative to private enterprises and Song et al. suggest strong connections between state-owned enterprises and state-owned banks (2011, pp. 202–204). The biological determinism assumed on the part of entrepreneurs inheriting skills perfectly is necessary to ensure that E- and F-firms differ in productivity. The fact that the managers are the children of

the owners could open for intergenerational altruism rather than the incentive approach to managerial compensation, so the framing in corporate governance is not necessary per se.

3.1.2 Equilibrium during transition

Due to the difficulties E-firms face in obtaining external financing, there will be a lower ratio of capital to output in E-firms.

LEMMA 1: *Let assumption 1 hold. Then E-firms will have a lower capital-output ratio and capital-labor ratio than F-firms.*

There are three further key properties of the model as it has been presented: F-firm capital follows from profit maximization, while E-firm capital and productivity are state variables; the level of capital per efficient unit of labor is constant for both E- and F-firms; and the saving of entrepreneurs in any period is linear in the current level of capital. The equilibrium dynamics of the model then follow:

LEMMA 2: *Given the level of capital in E-firms, K_{Et} , and general productivity, A_t , the equilibrium dynamics of total capital and employment in E-firms during transition are given as:*

$$\frac{K_{Et+1}}{K_{Et}} = (1 + \gamma_{K_E}) \wedge \frac{N_{Et+1}}{N_{Et}} = \frac{(1 + \gamma_{K_E})}{(1 + z)} \equiv 1 + v_E$$

There exists a critical value for the efficiency advantage of E-firms that must be exceeded for the employment share of E-firms to grow over time:

$$\chi > \hat{\chi} < \infty$$

For there to be a transition of labor from F- to E-firms, the discount factor on consumption, the ability of entrepreneurs to repay loans and the interest rate on loans must be sufficiently large, and the population and productivity growth sufficiently small.

Capital is implied to grow at a constant rate here, which hinges on assumptions of constant returns to capital in E-firms, and entrepreneurial saving and income proportional to E-firm profits. Income inequality between workers and entrepreneurs will be growing. There is also an implied sweet-spot for transition given by the managerial share of income: Compensation being too low makes for poor entrepreneurs and less investment, while compensation being too high makes for falling E-firm profitability.

GDP per capita is growing as long as $\chi > \hat{\chi}$ due to reallocation of labor from F- to E-firms.

3.1.3 Foreign surplus, savings and investment

The balance sheet of banks is given the standard way: Assets must equal liabilities. Given the definitions of capital accumulation, wage growth and population growth presented so far, there is a third lemma:

LEMMA 3: *The foreign surplus will be growing during reallocation of labor from F- to E-firms. Sufficiently credit constrained E-firms will make for a foreign surplus post transition as well.*

There will be less need for domestic financing of capital investment during transition, which means capital will migrate abroad. To achieve the particular dynamic, the credit constraint on E-firms must be sufficiently severe and the managerial share of income sufficiently small.

Reallocation of the labor force makes for a foreign surplus because the saving rate will increase and the investment rate will decrease during reallocation. Growth of aggregate savings is due to E-firm saving being higher than F-firm saving, which hinges on the entrepreneurial saving rate being higher than the saving rate of workers. For entrepreneurial savings to be higher one must assume that the intertemporal elasticity of substitution is larger than or equal to unity, $\theta \geq 1$. The decreasing investment rate follows from capital density being generally lower in E-firms, so any worker reallocated to an E-firm will work with less capital. Domestic investments will fall if one assumes no population and productivity growth.

3.1.4 Equilibrium post-transition and financial development

At the end of the transitional period all workers are employed in E-firms and the model collapses to a standard overlapping generations model. There are no F-firms, so the capital stock is given by the level of employment in E-firms. E-firms accumulate capital until the firm rate of return equals the domestic interest rate on loans, or until efficiency unit capital converges to a level that exceeds said interest rate. During convergence, efficiency unit wages and output as well as the foreign surplus are increasing as the rate of return on capital falls.

Due to the changing composition of firms, the investment rate is falling in the model as it has been presented here: Financially constrained firms expand, while unconstrained firms disappear. The investment rate in China has however not been decreasing over the proposed period of reallocation. Rather, the investment rate has been u-shaped (Bai et al., 2006, p. 69). To reconcile the model with u-shaped investments, Song et al. propose financial development during the period of transition. All else being equal, the capital-labor ratio and the wages in both E- and F-firms will increase during reallocation. The tendency of investment rates to fall is then offset. Increasing wages makes lower profits in E-firms and entrepreneurs will have to save more to be able to employ more workers. Both features make for a slower reallocation.

There were many reforms of the financial sector in China during the late 1990s and early 2000s which have had some effect on the lending behavior of Chinese banks (Podpiera, 2006). The reforms do not seem to have made a difference for the credit market access of private enterprises. The financial development is therefore more realistically conceptualized as a decreasing iceberg cost of financial intermediation than as an increasing share of loans to E-firms (Podpiera, 2006; Song et al., 2011, p. 204).

Here, then, is one way of getting some of the same dynamics from a model as observed for the Chinese economy. It is not the only way of modelling Chinese growth, but the benchmark calibration fits the data remarkably well.

4. Replication, methodology, data and conditions for the analysis

4.1 The original calibration

Growing Like China presents a deterministic rather than stochastic dynamic general equilibrium model (Song et al., 2011). Calibration is still needed and can indeed be problematic (Hansen & Heckman, 1996; Shoven & Whalley, 1992). That being said, we can expect the results of alternative calibrations to be tractable to model characteristics in a quite transparent way.

The model Song et al. use for their quantitative results is more or less the same as their theoretical model, but rather than assuming a two-period overlapping generations model, the quantitative model abstracts the lifespan of each agent to T periods (2011, pp. 220–221). The abstraction makes for some formal changes to the presentation of the budget constraint of the agents etc. The depreciation rate on capital is now assumed to be less than full for each period, $\delta < 1$. The iceberg cost of financial intermediation decreases over time so that the model ensures a u-shaped investment rate. The model is then programmed in MatLab to get what is presented as Figure 8 in Song et al. (2011, p. 223).

To solve the model numerically, calibration is needed. Song et al. calibrate a few parameters exogenously and establish the rest within the framework of the model (2011, p. 221). The focus of the calibration are the empirical moments of the period 1998-2005.

The exogenously calibrated parameters are (Song et al., 2011, pp. 221–222):

- *Entrance and exit of agents into the economy, including retirement age.* Agents enter the economy at age 28, retire at 58 – the average retirement age in China – and die at 78.
- *The annual deposit rate.* From 1998 until 2005 the average one-year CPI-deflated deposit rate was 1.75%.
- *The capital share of income.* “The capital share is set to $\alpha = 0.5$, consistent with Bai, Hsieh and Qian (2006)...”(Song et al., 2011, pp. 221–222). It is unfortunate that we have to guess what Song et al. mean by consistent. I understand it to be an approximation

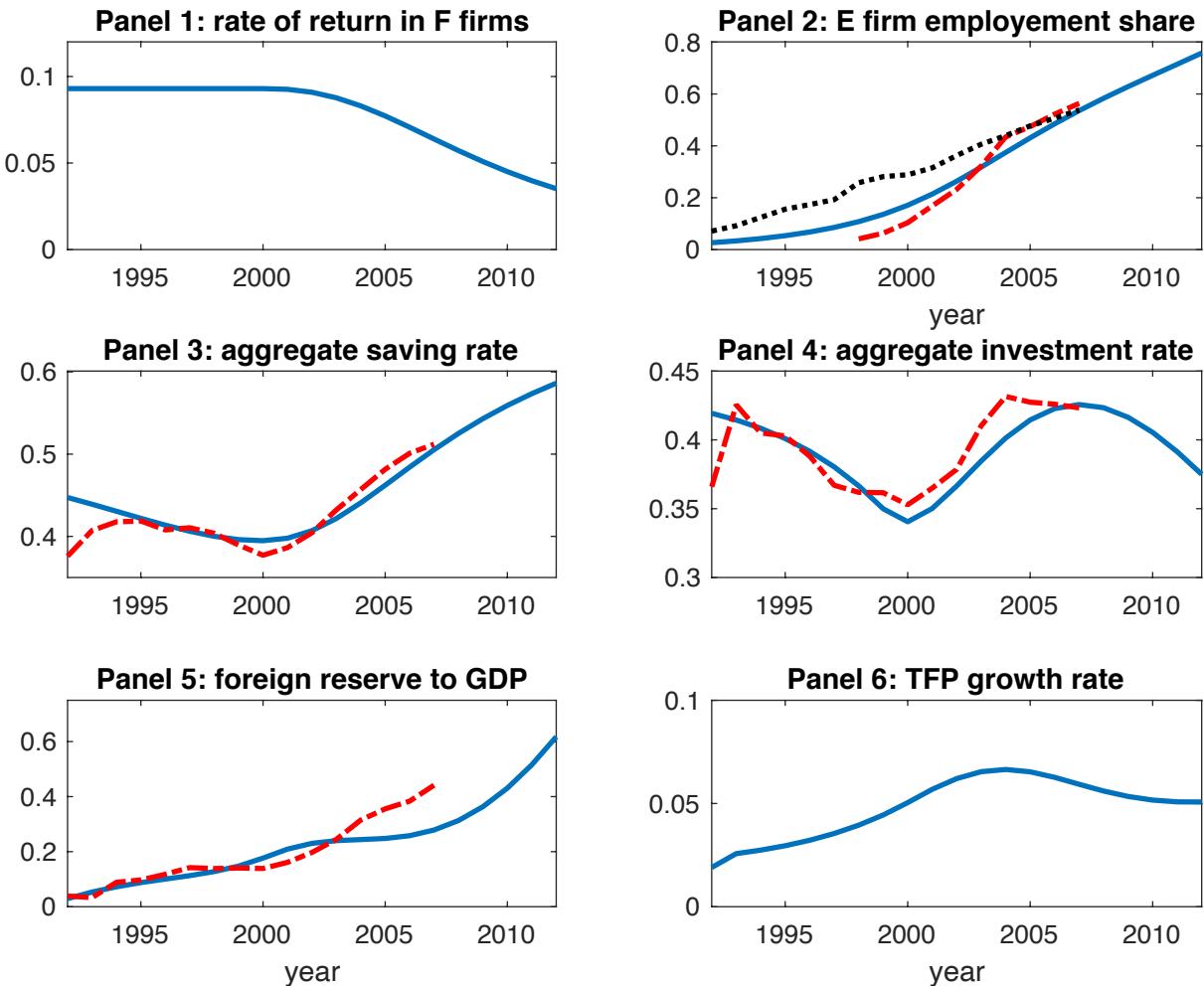
of the average capital share implied by the time series in Chong-En Bai, Chang Tai Hsieh and Yingyi Qian (2006, pp. 72–73), which is 0.5051.

- *The capital depreciation rate.* In the same sentence as the justification for the capital share, capital depreciation is established: “... and the annual depreciation rate of capital is set to $\delta = 0.1$. ”(Song et al., 2011, p. 222). Whether or not the depreciation rate is also meant to be consistent with Bai et al. is ambiguous: There is no explicit reference to the depreciation rate in Bai et al. but the average of the time series provided by Bai et al. would be approximately 0.1 (Bai et al., 2006, pp. 72–73). Elsewhere in Bai et al. the depreciation rate is referred to as assumed to be 8% for structures and 24% for machinery (2006, p. 70).
- *The annual population growth rate.* Consistent with the annual urban population growth during the period 1998-2005 as reported by the World Bank, the model population growth rate is set to 0.03.
- *The intertemporal elasticity of substitution:* $\theta = 2$, which means that the parameter for constant relative risk aversion is set to 0.5 in the MatLab code. There is no reference to any literature and no direct discussion of why the elasticity might be exactly 2. From the presentation of the theoretical model we know that Song et al. assume a lower limit for the elasticity to ensure a growing foreign surplus (2011, p. 216). Song et al. do a robustness test of their results where they assume people to have logarithmic preferences, $\theta \rightarrow 1$, as an alternative to their benchmark.

The specific calibration makes for a set of empirical moments that are both quantitatively and qualitatively similar to the observed time series for the relevant empirics from the Chinese economy: The private employment share, the aggregate saving rate, the aggregate investment rate, and the rate of the foreign reserve to GDP. Data for the private employment share are taken from measures of employment in manufacturing and in the urban sector. The share of private employment in manufacturing is estimated from reports by the National Bureau of Statistics in China (1998-2007), while the share of private employment for the urban sector is estimated from reported numbers in the China Labor Statistical Yearbook (1992-2007). Data on the investment rate are taken from various issues of the China Statistical Yearbook and the China Economy and Trade Statistical Yearbook. The saving rate is estimated as residual GDP accounting for private and state consumption. Finally, the data on the foreign reserve rate are taken from various issues of the China Statistical Yearbook. The comparison of the calibrated

model with the share of private employment and the saving, investment, and foreign reserve rates is presented in Figure 8 in Song et al., reproduced here as Figure 1.

Figure 1 – Replication of the benchmark calibration in Song et al. (2011, p. 223)



Note: Figure 8 presents the time series of six key variables: Rate of return in F firms, employment in E firms, the aggregate saving rate, the aggregate investment rate, foreign reserve to GDP, and productivity growth. The results of the calibration are the solid blue lines, while the red dashed lines are the observed time series. In panel 2 both aggregate and firm level employment are presented along with the calibration. The black dotted line is firm level employment, while the red dashed line is aggregate employment.

As argued by Song et al. (2011, p. 220), the benchmark calibration fits remarkably well with data, represented in Figure 1 by the red and black lines. The model reallocation of labor from state-owned to private firms follows the pace of observed reallocation on both the firm level and the aggregate level. During the reallocation productivity grows faster along with the pace of reallocation, in line with theories of misallocation and growth (Diao et al., 2017). The rate

of return in F-firms is decreasing along with a programmed decrease in the iceberg cost of financial intermediation. As the iceberg cost decreases, the saving rate increases along with the investment and foreign reserve rate. The u-shape of the investment rate reflects less investment during the initial reallocation of labor from F- to E-firms followed by an increase as the iceberg cost starts to decrease.

The calibration in Song et al. only partially fulfills the criteria for a calibration to be sound: The values used are not argued for in detail, but there are robustness tests (2011, pp. 225–227). The tests are: No financial development, no borrowing for entrepreneurs, low productivity advantage for E-firms, and logarithmic preferences. With the exception of the test assuming logarithmic preferences the robustness checks do not concern the parameters calibrated exogenously. Assuming there is no financial development makes for a faster transition as the rate of return to E-firms remains high. No entrepreneurial borrowing makes for a slower transition as entrepreneurs are more credit constrained. A lower productivity advantage makes for lower E-firm returns and a slower transition as entrepreneurial income is low. The case of lower productivity advantage is quantitatively identical to the case of logarithmic preferences according to Song et al. (2011, p. 227), which makes sense as logarithmic utility would make for lower entrepreneurial saving.

The robustness checks done by Song et al. are tests of model features rather than robustness to measurement error or researcher bias (Glaeser, 2006; Leamer, 1983). Excluding the test of logarithmic preferences there is a tacit assumption that all values used for calibration are correct. Such an assumption is unwarranted when the exogenous parameters are not argued for in a detailed way. One would expect some argument as to why the capital share of income is 0.5 other than postulated consistency with one study⁵, why the depreciation rate is 0.1, or whether there are any empirical grounds for holding the intertemporal elasticity of substitution to be 2.

The concern with calibration is in large part that the results are reflective of the parameters that are calibrated exogenously rather than actual model behavior (Blanchard, 2016; Canova & Sala, 2009). From a critical perspective, then, it would make sense to extend the results by altering the values of the exogenously calibrated parameters to check for robustness. The capital share

⁵ For instance, Bai et al. discuss the concern for measurement error in the labor share of income (2006, pp. 71–74).

of production, the depreciation rate, and the intertemporal elasticity of substitution will be the focus of my robustness checks. The other parameters that are calibrated exogenously – the time at which agents enter, retire from and exit the economy; the annual deposit rate; and the annual population growth rate – seem less controversial by the nature of measurement. In so far as the estimates are averages, the nature of what is averaged does not give room for much measurement error.

4.2 Methodology

The alternative calibrations are based on research relevant to the calibrated parameters. Tests using relevant data arguably falls within the methodology of calibration as suggested by Kydland and Prescott (Kydland & Prescott, 1991, 1996), while taking the critique of calibration seriously. The values used to extend the original calibration are subject to the same criteria as laid out for sound calibration: The research the estimates are based on must be presented, the relevance of the estimates for the model must be argued for, and if the alternative values violate the model assumptions in any way one must do an extension that excludes the assumption violating values.

The replication itself is done in MatLab as used by Song et al. The original code is copied and elaborated on in separate scripts. There were no scripts for the original robustness checks in the appendix to *Growing Like China*, so I made my own based on the scripts from another article featuring calibration (Jones & Klenow, 2016). My versions of the original scripts for the model and the scripts for the extensions are available as a separate appendix along with the thesis. I include one of my scripts for the alternative calibrations as an example in appendix B.

I use the following method for the model extensions: Each alternative estimate of a parameter is used as the basis for a new calibration, alone and in any possible combination with other alternative estimates. I calibrate the model for an alternative capital share and proceed to combine that specific alternative capital share with all estimates of the capital depreciation rate and intertemporal elasticity of substitution, including the benchmark values. By reviewing relevant literature, I have found seven alternative estimates of the capital share, five alternative estimates of the depreciation rate, and two alternative measures of the elasticity. Including the benchmark values for the three parameters makes for a total of 168 calibrations.

The robustness of the extensions is evaluated in two ways: By excluding all alternative parameter estimates that are inconsistent with the model assumptions made by Song et al., and by varying the capital share, depreciation rate, and elasticity separately and incrementally from the benchmark values. The incremental changes are done for twenty percent intervals using the benchmark value as a mean: Ten percent above the benchmark value and ten percent below the benchmark value. That makes for twenty alternative calibrations for each of the three parameters.

The actual evaluation of robustness will be visual, similarly to the evaluation done by Song et al. Visual evaluation of model fit⁶ is also in line with calibration as practiced by Kydland and Prescott originally (1982). Evaluating model results visually is somewhat problematic and has been criticized extensively (Andersen, 1991; Hansen & Heckman, 1996; Sims, 1996). There are proposed methods for determining the fit of a calibrated model in a less ad hoc manner (DeJong, 2011, pp. 279–284). Although such measures would be relevant for my purpose, time and the scope of the thesis speaks to leaving the measures of fit for later research.

I am of course also subject to biases in my attempt to evaluate calibration and like any other researcher I am hard pressed to be bias-free in my conclusions. The thesis must not be seen as an attempt to disregard calibration completely or, for that matter, at all. Calibration can be well-founded. What my thesis is, is a limited enquiry into the robustness of a specific case of calibration, based on estimates taken from a limited number of articles deemed relevant by me.

4.3 Estimates used for extensions

4.3.1 *The capital share of income in China*

I am working from the assumption that the benchmark calibration in Song et al. uses a capital share of output that is an approximation of the sample average of the capital share time series presented in Bai et al. (Bai et al., 2006, pp. 72–73; Song et al., 2011, pp. 221–222). A presentation of the paper by Bai et al. is therefore in place. The usual way of computing the

⁶ As Torben Andersen has argued, to speak of “a good fit” and disregard statistical testing as irrelevant seems inconsistent (1991, p. 180).

capital share is as a residual of the labor share. Residual measurement is also the method of Bai et al. They have access to annual data on the labor share on the provincial level from the National Bureau of Statistics, which they transform to economywide labor shares by averaging provincial labor shares weighted for the respective share of GDP of each province.

There are two general concerns about the data: The labor share may be underestimated due to unmeasured nonwage benefits or because the reported numbers exclude labor income of the self-employed (2006, pp. 71–74). Bai et al. dismiss the first concern because the National Bureau of Statistics explicitly includes nonwage benefits in their labor share measures. The second concern, they argue, is potentially inverted, as the bureau counted all self-employment income as labor income until 2005, which could lead to overestimation. The capital share is by no means measured exactly and we may be concerned about some underestimation in the series provided by Bai et al. (2006, p. 74).

The problem of measuring the labor share of output has been extensively documented by Marta Guerriero in a recent paper (2019). She presents six alternative measures of the labor share – one of which is proposed as a new measure – all with their own strengths and weaknesses. Guerriero provides estimates of China's labor share for two out of the six measures.

The most common measure is what Guerriero calls LS1 or the unadjusted labor share: Total compensation of employees to net value added (2019, p. 3). The unadjusted share is what Bai et al. base their estimates on, and there may be reasons to fear underestimation. Guerriero estimates the unadjusted share to be 0.54 for China for the period 1970-2015 (2019, p. 18).

The other measure Guerriero estimates for China, LS5, is somewhat similar to the unadjusted labor share but total compensation is weighted by the fraction of total workers to number of employees (2019, p. 5). Guerriero argues that the weighted share is a better measure as it accounts for a changing workforce composition and may be superior in economies where many are self-employed. A potential problem with the measure, other than access to all the relevant data for estimation, is that it is sensitive to systematic income differences between normal employees and the self-employed. The weighted share is estimated to be 0.59 for China for the period 1970-2015 (2019, p. 18).

The work of Guerriero implies capital shares below what is assumed in the calibration by Song et al. and what one may infer from averaging the time series of Bai et al. over the period 1978–2005. The discrepancy could be because the benchmark calibration is supposed to match a shorter time span than what the unadjusted and weighted labor shares are calculated for in Guerriero. There are, however, no justifications for the use of averages over shorter time spans in the calibration literature. Kydland and Prescott emphasize a match between the time horizon of the statistics computed from the model and the observed statistics (1996, p. 75), but that does not automatically exclude using averages for a longer time span for calibration. That leaves the concern of breaks in the time series averaged. In so far as the estimates of the capital share presented in Bai et al. are indicative of the actual series, there are no signs of breaks (2006, pp. 72–73). The model makes very clear predictions about transitions and the steady-state as well. One would have reason to expect the model to be robust to calibrations using long-term averages⁷ (Hansen & Heckman, 1996, p. 95). We might still be concerned that newer data biases the estimate downwards in Guerriero due to the ten years from 2005 to 2015. It is reasonable, then, to test the quantitative predictions for estimates of the capital share that were done some time before the paper by Song et al.

Chang-Tai Hsieh and Peter Klenow (2009, p. 1404) have done a study of the same theory of misallocation in China as in Song et al., but their main endeavor is empirical rather than to build a new model. The labor share is estimated to have a median value of 0.3 at the plant level of production, and is assumed to be 0.5 at the aggregate level (2009, p. 1413). The choice of capital share by Song et al. gets some confirmation from Hsieh and Klenow, even if it is principally through another assumption. The discrepancy between the measured and the assumed capital share in Hsieh and Klenow may be typical: It is a well-known problem that micro- and macro-level often do not give estimates that are consistent with one another (Hansen & Heckman, 1996; Simon & Aigner, 1970). The method of calibration in itself is however built on the assumption that microfounded models can use both micro and macro data for their solutions (Prescott, 1986, p. 22). I think it reasonable, for the sake of testing for what the upper bounds of the capital share in China might be, to use the plant level data from Hsieh and Klenow as well.

⁷ The model is indeed a long-term model and so the concern about short-term dynamics not being correctly identified by long-term averages other than in special cases should not apply in any direct sense (Hansen & Heckman, 1996, p. 95).

Alwyn Young's research on China is related to the theory of misallocation: If one accounts for higher labor force participation, reallocation of labor out of agriculture, and higher educational attainment, the growth rates experienced by China from 1978 to 1997 are reduced to levels similar to those experienced by other growth economies at other points in history (2003). The Chinese growth experience can be explained well by an augmented neoclassical growth model once factor accumulation is accounted for. The labor share in Young's paper is estimated as part of unmodified GDP and as part of GDP excluding agriculture⁸ (2003, pp. 1254–1258). The estimates are the same as those reported by the National Bureau of Statistics, the unadjusted labor share in Guerriero, that is (2019, p. 3). Young agrees with Bai et al. that the labor share in China does not need adjustment: The self-employed are relatively few and are explicitly counted by the bureau as receiving labor income (2003, p. 1257). The average labor shares for the period 1978-1997 are 0.6 for GDP and 0.46 for GDP excluding the agricultural sector (2003, p. 1255).

The reader will probably wonder if my use of the estimate of the weighted labor share provided by Guerriero (2019, p. 18) is well-founded. Other authors are convinced that the unadjusted labor share is a correct estimate in the case of China (Bai et al., 2006; Young, 2003). Although there might be reasonable objections to an adjusted estimate, I refer to the correlation matrix of the differing labor shares provided by Guerriero (2019, p. 10) which shows a correlation of 0.554 between the estimators for the unadjusted and weighted labor shares. I think it safe to conclude that the weighted share accounts for some variation not captured by unadjusted share and so testing for the weighted share should be justified.

A final measure of the capital share I found in a paper by Gregory Chow. He investigates the effect of capital accumulation on Chinese growth from the 1950s to the 1980s (1993). Chow's investigation seeks to uncover the effect of the particular emphasis on capital accumulation in the Chinese economy and whether the disturbances of the Great Leap Forward and Cultural Revolution have had any effect on growth. Much like Young (2003), Chow finds that a standard neoclassical growth model can account well for the Chinese growth experience up until the 1980s⁹ (1993, p. 841). The political events of the period are also found to have had detrimental

⁸ By excluding agriculture, the reallocation of labor to the nonagricultural sector becomes an input (Young, 2003).

⁹ Chow assumes no technological growth in China during the period of study so that the only source of growth is capital accumulation.

effects on aggregate and sectorial growth. Chow argues for a capital share of 0.6 in China during the period of study (1993, p. 826): The choice fits the data well, it implies absence of technological progress, and it is consistent with cross-country results for the same period provided by Mankiw, Romer, and Weil (1992, pp. 414–415). The capital share being as large as 0.6 may be a relic of the past and not necessarily representative of the reform period in Song et al. The capital share in Chow is however lower than what Hsieh and Klenow find at the plant level (2009, p. 1413), and so it may be a reasonable value to test the model for as it lies between the original calibration and the share found by Hsieh and Klenow.

Indeed, all of the alternative capital shares I have found may be representative of a longer view of the Chinese economy or another period in the evolution of the Chinese economy. One could argue that the model of Song et al. cannot be falsified on a quantitative basis by observations that fall outside the period they calibrate their model for as it is a special period. It would however be a shame if the model is not robust to a longer view of growth factors in the Chinese economy, even if the goal of the calibration is to explain a seven-year period. Needless to say, I will test for all the alternative capital shares I have found.

4.3.2 The capital depreciation rate in China

The capital depreciation rate is notoriously hard to pin down, because it is seldom estimated for specific countries and often merely assumed to be of a certain size when used in models (Schündeln, 2013, pp. 277–280). The value of the depreciation rate is also assumed in the case of the model in *Growing Like China* (2011, pp. 221–222). The assumption may or may not be consciously consistent with Bai et al., as the sample average implied by the statistics Bai et al. provide is approximately 0.1 (2006, pp. 72–73).

More or less all the alternative depreciation rates found for my thesis are based in assumptions rather than measurement. Most of the research I could find on China which is explicit about the depreciation rate either assume a size (Bai et al., 2006; Hsieh & Klenow, 2009; Perkins, 1988) or lean on the assumptions of other researchers (Wang & Yao, 2003). The actual size of the depreciation rate may be quite controversial, then, and it should be reasonable to use the educated guesses of other researchers for alternative calibrations (Schündeln, 2013; Sims, 1996). The problem of pinning down the depreciation rate is underlined by what the estimates in Bai et al. are based on: They assume a depreciation rate of 0.08 for structures and 0.24 for

machinery (2006, p. 70). On the basis of the assumed lower and upper bounds, Bai et al. generate a times series for the depreciation rate with an implied average of about 0.1 (Bai et al., 2006, pp. 72–73). In other words, the span of reasonable values for the depreciation rate is considerable.

A broader survey of the literature turns up estimates of the depreciation rate that are mostly much lower than the benchmark of 0.1. The research I refer to here is not necessarily about the transitional period 1998-2005, but that may not be detrimental for the purpose of testing: The depreciation rate is identical in three papers published over the time span 1988-2009.

First, the paper on misallocation in the Chinese economy by Hsieh and Klenow reports results based on an assumed depreciation rate of 0.05 (2009, p. 1414). Hsieh and Klenow do not justify the assumption in any explicit way. Second, in their paper on Chinese growth, Yan Wang and Yudong Yao assume a depreciation rate of 0.05 (2003, p. 37) with reference to another paper by Dwight Perkins (1988). The third paper using a depreciation rate of 0.05 is therefore Perkins's, and it is again an assumed value (1988, p. 629). There is no better basis for a depreciation rate of 0.05 than informed assumption and assumptions based on the assumptions of other authors.

Wang and Yao subject their results to robustness tests by changing the depreciation rate. Depreciation is alternatively assumed to be 0.1 and 0.15 (2003, p. 37). Their estimated growth rate seems quite sensitive to alternative depreciation rate for what they call the pre-reform period, 1953-1977, while the estimates for the whole sample, 1953-1999, and for the reform period, 1978-1999, are seemingly robust. As the alternative depreciation rate of 0.1 is identical with the benchmark calibration in Song et al., a test would not make any sense. I only test the model using Wang and Yao's alternative rate of 0.15.

The final depreciation rate used for my tests I found in a report by the Organization for Economic Co-operation and Development (OECD, 2000). The report is concerned with the methodology for national accounts in China. The depreciation of fixed assets is reported as 0.04 for the sectors investigated: Real estate, government, and other services (2000, pp. 42–46). The estimate of 0.04 applies to newly created fixed assets, which is reported in the *Annual Report on Fixed-Asset Investment* by the National Bureau of Statistics. There is some concern about underestimation due to the accounting methods of the bureau: They do not consider the high

inflation of the reform period from 1978 onwards (2000, p. 16). We may have good reason to be skeptical about such a low depreciation rate, but note also how close the estimate is to the assumed depreciation rate of other papers (Hsieh & Klenow, 2009; Perkins, 1988; Wang & Yao, 2003).

When it comes to the depreciation rate there seems to be no better estimates than the educated guesses of other researchers. The depreciation rate assumed by Song et al. could be as well-founded as any other estimate that can be found in the literature. There is still a case for checking the model for depreciation rates assumed by other researchers, given the variation in the size of the estimates that can be found for China specifically. In so far as expert opinion differs on the size of the depreciation rate, we will do well to test the model for differing opinions (Blanchard, 2016; Sims, 1996). Using assumed parameter sizes from other research is in line with how calibration is practiced (Canova & Sala, 2009, p. 435).

4.3.3 The conundrum of the estimated intertemporal elasticity of substitution

The problem of measuring the intertemporal elasticity of substitution is well-known. A classic study by Robert Hall (1988) suggested an elasticity close to zero for aggregate consumption series, which in turn spurred a wide research program to find reasons as to why the elasticity does not show up in data.

The rejection of intertemporal elasticity of substitution is more prominent with authors using aggregate time series (Campbell & Mankiw, 1989; Mankiw et al., 1985). Studies of aggregate consumption are however subject to critique due to possible measurement and specification errors. Aggregation over many population sub-groups may hide the optimizing behavior of people over their life-cycle due to differences between people, non-linearities, and entries and exits to the economy (Attanasio & Weber, 1993, p. 636). Studies that use food consumption as a proxy for all consumption when estimating the elasticity will have a specification error, as such studies do not account for the effect of the interest rate on durable consumption (Attanasio & Weber, 1993, 1995; Ogaki & Reinhart, 1998). As the interest rate is more likely to affect durable consumption, estimates of the elasticity based on non-durable consumption will be biased downwards (Ogaki & Reinhart, 1998, pp. 1079–1080). Another possible specification error is that nondurable consumption may be affected by predictable changes in household composition and household labor supply over time. Such time dependent change ought to be

accounted for, or the results may point to excess sensitivity of consumption to growth in labor income (Attanasio & Weber, 1995, p. 1123). Studies based on household data usually estimate an elasticity significantly different from zero and considerable difference in behavior across households, countries etc. (Atkeson & Ogaki, 1996; Attanasio & Weber, 1993, 1995; Biederman & Goenner, 2008; Ogaki & Atkeson, 1997; Ogaki & Reinhart, 1998; Okubo, 2011).

The inconsistency between aggregate and household level estimates of the intertemporal elasticity of substitution is somewhat at odds with the original justifying assumption for calibration: “A fundamental thesis of this line of inquiry is that the measures obtained from aggregate series and those from individual panel data must be consistent. After all, the former are just the aggregates of the latter.” (Prescott, 1986, p. 22; see also Tobin, 1950).

The ambiguity about the size of the elasticity alone cannot justify that a calibration is based on some assumed value. One would at least expect some theoretical justification as to why the elasticity is of a certain size. In the model of Song et al. there is one such justification. Savings should be exclusively non-decreasing in the rate of return (2011, p. 208) and Song et al. wants to ensure higher saving by entrepreneurs than by workers (2011, p. 216). Based on the lower-bound assumption, an initial benchmark guess of an elasticity of 2 may be quite reasonable and Song et al. do test their results for a different elasticity (2011, pp. 225–226). The elasticity being assumed is in line with the actual practice of calibration (Canova & Sala, 2009, p. 435) and some of my extensions are arguably using assumed values as well, even if I am not the originator of the assumptions (Hsieh & Klenow, 2009; Perkins, 1988; Wang & Yao, 2003). But then, what is the point of calibration and how do the models based on calibration differ from the supposedly inferior models of the 1970s (Romer, 2016)? If results are to be credible, one would at least expect some adherence to the evidence of the magnitude of a parameter.

What does the evidence on the intertemporal elasticity of substitution say? Based on the studies of household data the elasticity does exist but is not very large, somewhere between 0 and 1. The elasticity being quite small is the case for studies that allow for different behavior between households (Atkeson & Ogaki, 1996), substitution between durable and nondurable goods (Ogaki & Reinhart, 1998) and predictable changes in household composition (Attanasio & Weber, 1995). As one would expect, there is considerable difference between countries but it is striking that most statistically significant results suggests an elasticity between 0 and 1 (Havránek et al., 2015, p. 112). The fact holds for individual country comparisons (Atkeson &

Ogaki, 1996; Okubo, 2011) and for a larger set of countries (Havránek et al., 2015)¹⁰. All evidence points towards a smaller elasticity than what Song et al. use in their model. I will therefore do tests using some relevant estimates of the elasticity. The use of the estimated elasticity may not be consistent with the theoretical framework of the model (2011, p. 202).

4.3.4 The estimated elasticity using data from China and “best practice” estimates

The tests of the model subject to alternative intertemporal elasticities will be based on a meta-analysis done by Tomáš Havránek (2015) and a cross-country study Tomáš Havránek, Roman Horvath, Zuzana Irsova, and Marek Rusnak (2015). The study by Havránek et al. I deem relevant for the simple reason that it includes the estimated elasticity for China and that the estimate is statistically significant at a 5% level (Havránek et al., 2015, p. 112), The study by Havránek is interesting due to its focus on bias in reported results (Havránek, 2015, pp. 1183–1188).

The cross-country meta-analysis by Havránek et al. aims to investigate the degree of country differences for the intertemporal elasticity of substitution. Havránek et al. collect and replicate a large number of individual country studies and use Bayesian Model Averaging to obtain an average estimate (2015, p. 205). The average elasticity in the study is 0.5, with considerable variation across countries. Exactly what drives the variation is not clear, but the level of wealth and financial market participation seem important (2015, pp. 107–108). It is worth noting that 80% of the studies included in the meta-analysis are based on aggregate and long-run time series, which may bias the estimated elasticity downward according to the literature (Atkeson & Ogaki, 1996; Attanasio & Weber, 1993, 1995; Biederman & Goenner, 2008; Ogaki & Atkeson, 1997; Ogaki & Reinhart, 1998; Okubo, 2011). Havránek et al. argue that the biasing effect in their meta-analysis can be ambiguous due to how different studies are designed, and they test their estimates for design difference (2015, p. 103). The estimated average elasticity for China is close to the average for the whole sample: 0.530 (2015, p. 112). The estimate is based on five individual studies, and yields a standard error of 0.234, which implies significance at the 5% level if one assumes a normal distribution as in the other meta-study by Havránek (2015, p. 1186).

¹⁰ There are three notable exceptions where the estimates are significant and quite large: Ireland, Kenya and New Zealand (2015, p. 112).

Havránek (2015) has done a meta-study of papers on the intertemporal elasticity of substitution where the aim is to check for any significant effect of reporting bias. There is strong evidence of selective reporting, meaning researchers discard negative or insignificant point estimates too often. The result is an upward bias in reported estimates of the elasticity which are comparable in magnitude to the difference between estimates of the elasticity using micro data and estimates of the elasticity using macro data. Havránek produces an estimate for the elasticity with 95% confidence bounds using what he deems to be the best practice of studies estimating the elasticity and controlling for the reporting bias¹¹. The point estimate using the best practice method is 0.36 with a 95% confidence interval of [-0.2, 0.8] (2015, p. 1196).

The elasticity may very well be zero even for micro data once the reporting bias is accounted for. Havránek concludes that calibrations using a higher elasticity than 0.8 is inconsistent with the bulk of the empirical literature (2015, p. 1196). It seems reasonable, then, to test the present model for the point estimate reported by Havránek, even if the estimates do not come from Chinese data per se.

As we would have grounds to suspect from the general discussion of the estimated intertemporal elasticity of substitution, the estimates I have presented here are all lower than the what the theoretical assumption about the elasticity in Song et al. allows for (2011, p. 216). A number of alternative calibrations will arguably be invalid on the basis of theoretical considerations. Also, there may very well be other assumptions that are violated by the use of alternative capital shares or depreciation rates.

4.4 Model assumptions and violations

If there is to be an equilibrium with E-firms in the model one must ensure that entrepreneurs choose delegation as their mode of corporate governance. The efficiency gains due to delegation to a manager, χ , must therefore be larger than the loss of profits due to the managerial share of income, ψ .

¹¹ Havránek's best practice is admittedly subjective, but it does account for all the issues regarding the estimation of the elasticity mentioned above: One must take account of both durable and non-durable consumption, micro data allows for more behavioral difference between households etc. See Havránek's list (2015, p. 1196).

Assumption 1 then holds that:

$$\chi > \underline{\chi} \equiv \left(\frac{1}{1-\psi} \right)^{\frac{1}{1-\alpha}} \Leftrightarrow \chi - \left(\frac{1}{1-\psi} \right)^{\frac{1}{1-\alpha}} > 0$$

If Assumption 1 holds, there will be a transition of the labor force towards E-firms as delegation will be chosen over centralization and young entrepreneurs will prefer reinvestment in E-firms over saving. The labor share features in Assumption 1 and also plays a role in the calculation of the efficiency advantage. The efficiency advantage must therefore be recalculated for every alternative capital share used, and Assumption 1 must be reevaluated.

What I will refer to as Assumption 2 follows from Lemma 2:

There exists a critical value for the efficiency advantage of E-firms that must be exceeded for the employment share of E-firms to grow over time:

$$\chi > \hat{\chi} < \infty$$

The calculation of whether or not Assumption 2 holds follows from what I have presented formally in the appendix as equation (11):

$$\begin{aligned} & \frac{\psi}{1-\psi} \frac{1}{\alpha(1+\nu)(1+z)} - \frac{1}{R^l} \left[(1-\psi)^{-\frac{1}{\alpha}} \chi^{-\frac{1-\alpha}{\alpha}} - \eta \right] \\ & - \beta^{-\theta} (1-\eta)^{1-\theta} \left\{ \frac{1}{R^l} \left[(1-\psi)^{-\frac{1}{\alpha}} \chi^{-\frac{1-\alpha}{\alpha}} - \eta \right] \right\}^\theta > 0 \end{aligned}$$

where ν is the population growth rate, z is the technological growth rate, R^l is the economywide interest rate on loans, η is the amount E-firms can borrow, and β is the discount factor on future consumption. $\hat{\chi}$ would be the efficiency advantage that equates the equation to zero.

The model also needs an assumption to achieve a growing foreign surplus. Assumption 3 is really three assumptions in one: The intertemporal elasticity of substitution must be larger than or equal to one; the managerial compensation must be sufficiently small; and the credit constraint on E-firms sufficiently severe. To evaluate Assumption 3, I use equation (14) in the appendix, reformulated as:

$$\frac{\alpha(1+z)(1+\nu)}{(1-\alpha)} \frac{1+\beta^{-\theta}R^{1-\theta}}{R^l} - \frac{\psi}{1-\eta(1-\psi)} > 0, \quad \theta \geq 1$$

Assumptions 1-3 make clear three characteristics of the model that deserve mention. First, a larger capital share will make transition slower per Assumptions 1 and 2. The feature makes intuitive sense, as more capital-intensive production will benefit F-firms rather than E-firms. Second, the managerial share of income cannot be too small nor too large if the model dynamics are to be in line with the theory proposed by Song et al. A share that is too small makes for poor managers and insufficient E-firm investment, while a share that is too large make E-firms unprofitable. Determinants of the managerial share of income will therefore play a role in whether or not the model achieves the wanted dynamics. One such determinant is the capital depreciation rate: A small depreciation rate makes for smaller managerial share of income as capital intensive production is more profitable and vice versa. Third, if the foreign reserve dynamic is to be achieved one needs to assure that the entrepreneurial saving rate is sufficiently larger than the saving rate of workers which depends directly on the intertemporal elasticity of substitution being larger than or equal to one.

We know, without calculating any assumption violations, that there will be quite a few of the alternative calibrations that are in direct violation of Assumption 3: All of the alternative calibrations using the estimated elasticities from Havránek (2015) and Havránek et al. (2015). One can argue that the calibrations using estimated elasticities are irrelevant on theoretical grounds. Still, how the model behaves when subjected to actual estimates of the elasticity should be of interest and the results of calibrations using the estimates will be presented regardless. I will include a robustness test where all calibrations that violate one assumption or more are excluded. The estimates used for alternative calibrations and whether these are in violation of any assumptions are presented in Table 1.

Table 1 – Estimates used and assumptions violated

Author(s)	Capital share	Capital depreciation	Elasticity of substitution	Assumption 1 violation	Assumption 2 violation	Assumption 3 violation
Bai et al. average	0.5051 (0.0049)			No	No	No
Guerriero unadjusted	0.46			No	No	No
Guerriero weighted	0.41			No	No	No
Hsieh & Klenow	0.7			No	No	No
Young aggregate	0.4			No	No	No
Young non-agricultural	0.54			No	No	No
Chow	0.6			No	No	No
Bai et al. mean		0.1082 (0.00082)		No	No	No
Bai et al. lower		0.08		No	No	No
Bai et al. upper		0.24		No	No	No
OECD		0.04		No	Yes	No
Hsieh & Klenow; Wang & Yao; Perkins		0.05		No	Yes	No
Wang & Yao upper		0.15		No	No	No
Havránek et al.			0.530	No	No	Yes
Havránek			0.33	No	No	Yes

Note: The rows refer to the studies that provide estimates of the capital share, the depreciation rate, and the intertemporal elasticity of substitution. The columns present the exact estimates of each study surveyed, and provide an overview of whether a particular estimate, alone or in combination with other estimates, violate any of the three assumptions I have presented. I have chosen to name the average capital share implied by Bai et al. as “Bai et al. average”, while the average capital depreciation rate implied by Bai et al. is named “Bai et al. mean”. (2006, pp. 72–73). Standard errors for the estimated averages are in parentheses.

5. Analysis

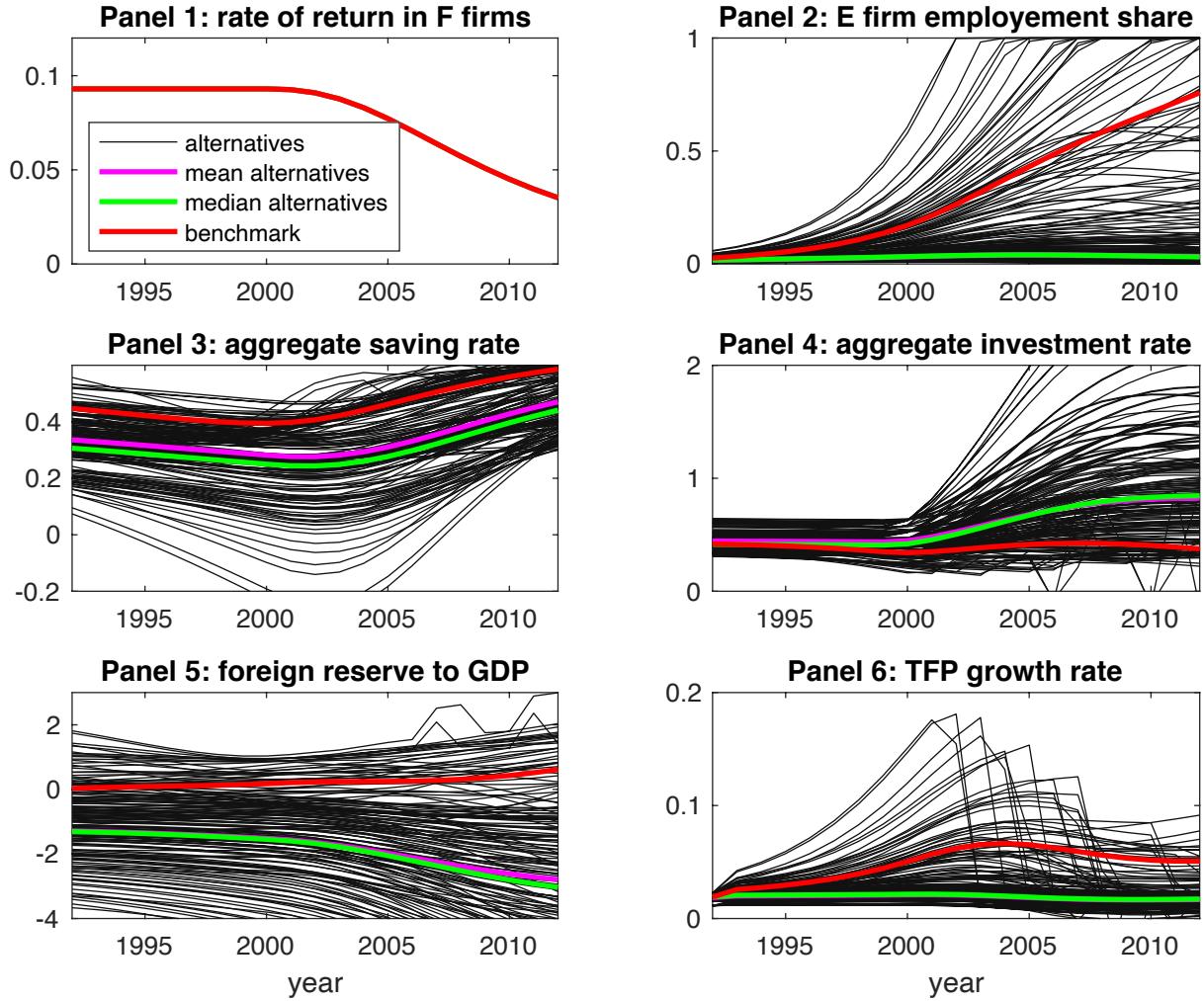
5.1 Extensions and robustness checks

5.1.1 *The model subject to all possible calibrations given alternative estimates*

Figure 2 presents a total of 168 unique calibrations, including the benchmark calibration in Song et al. (2011, p. 223), based on the alternative estimates of the capital share, depreciation rate and intertemporal elasticity of substitution. Note that the rate of return in F firms depends solely on the programmed decrease of the iceberg cost of financial intermediation which in turn is not dependent on any of the parameters altered in my tests. The time path in Panel 1 will therefore look identical for all the figures.

Looking at Panels 2-6, the variance due to alternative initial values of the capital share, depreciation rate, and intertemporal elasticity is quite extreme. Most alternative calibrations make for almost no reallocation of labor, which means almost no productivity growth. There are close to no E-firms at the mean or median level, which makes for generally lower saving rates. The mean and median saving rates are lower than the benchmark but follow a qualitatively similar time path. The mean and median investment rates are also qualitatively similar to the benchmark, but once there is financial development, represented by Panel 1, the bulk of alternative investment rates increase to very high levels. The u-shape of the investment rate time path seems to disappear for the mean and median alternative rates. Given the generally lower level of most alternative saving rates and that most alternative investment rates are initially quite similar to the benchmark, the mean and median foreign reserve rates are generally at a lower level than the benchmark. Once the alternative investment rates start to diverge from the benchmark, the alternative foreign reserve rates decrease rapidly. The test using 168 alternative calibrations makes for no reallocation, no productivity growth, an investment rate that is not u-shaped, and an increasing foreign reserve deficit rather than an increasing foreign reserve surplus. All results are counter to the theory of Song et al.

Figure 2 – Calibrations using all possible combinations of alternative estimates



Note: ‘alternatives’ refers to all the possible combinations of the alternative estimates of the capital share, depreciation rate and intertemporal elasticity of substitution, represented here by the black lines. The total number of black lines is thus 167. ‘mean alternatives’ is a calibration using the mean value of all possible alternative calibrations here represented by the magenta line. Similarly, ‘median alternatives’ is a calibration using the median value of all possible alternative calibrations, represented by the green line. The red line is the replicated benchmark calibration as presented in Figure 1.

Quite a few of the alternative calibrations in Figure 2 use an intertemporal elasticity that is smaller than one, which could help explain the generally lower level of the mean and median saving rate and by extension the foreign reserve deficit and lack of reallocation to E-firms. As constraints on borrowing disappear, F-firms borrow to invest in more capital while the saving rate remains low, and so the foreign reserve deficit increases. From Table 1 we know that the depreciation rate can be too low for Assumption 2 to hold and for the employment share in E-firms to grow. The low depreciation rate makes for a low managerial share of income and too

little reinvestment in E-firms. On the other hand, some of the higher depreciation rates would make for higher investment rates as capital must be replaced more frequently. Higher depreciation would also make for larger managerial shares of income and lower profitability of E-firms. The depreciation rate being high can therefore slow down transition and lead to slower productivity growth. The capital share being low would make for faster reallocation of labor and a lower investment rate, while a higher capital share would make for slower reallocation and a higher investment rate. A lower saving rate – due to, say, a smaller intertemporal elasticity of substitution – combined with a higher capital share could make for a foreign reserve deficit, the saving rate being less volatile compared to the investment and foreign reserve rate.

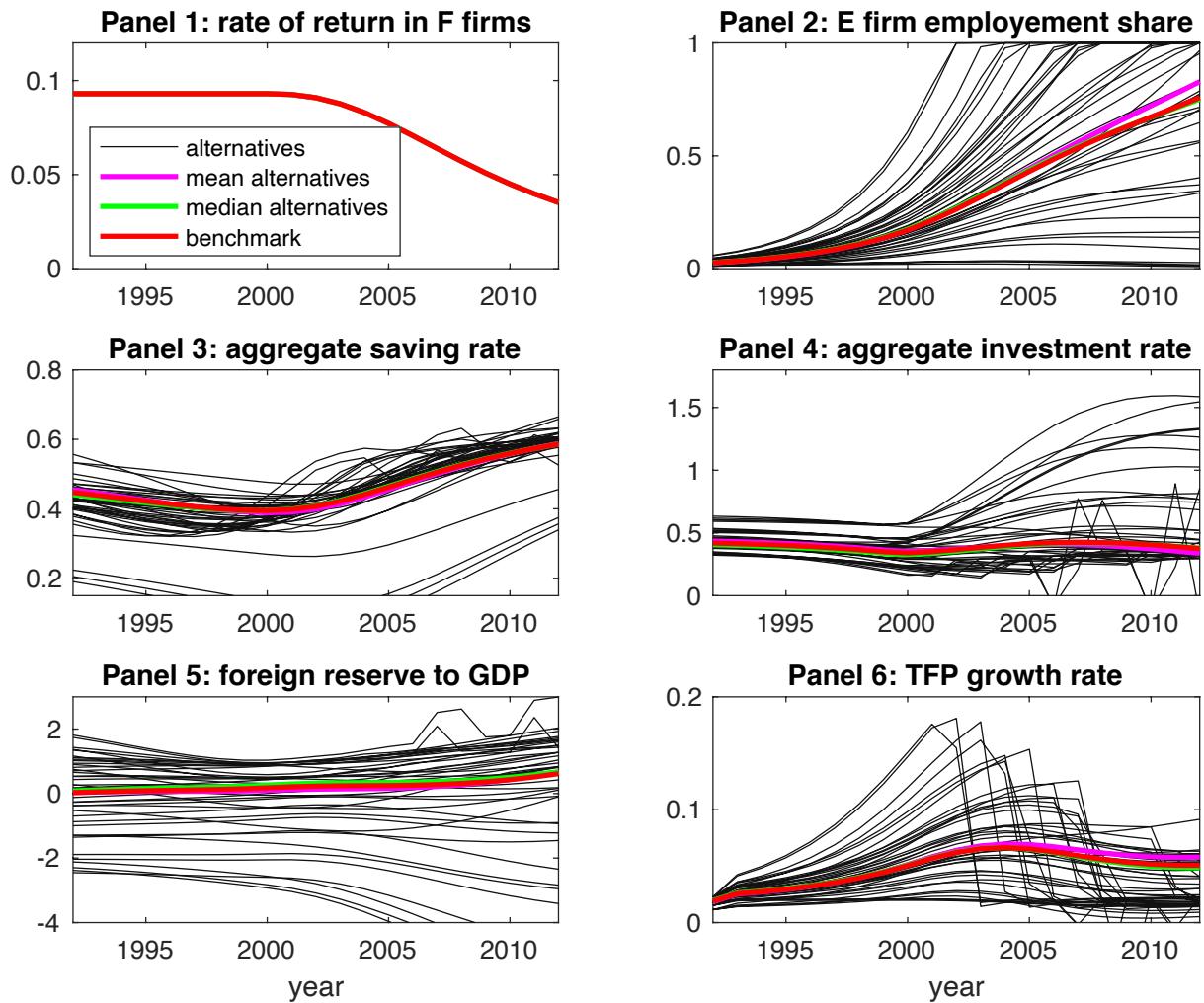
5.1.2 The model subject to assumption-consistent alternative calibrations

Some of the more extreme dynamics should disappear once the alternative values that violate the model assumptions are excluded. Figure 3 presents such an assumption-consistent test of the model.

The results in Figure 3 are quite different to those in Figure 2. The mean and median of the alternative calibrations are almost identical to the benchmark. Reallocation is a little faster on average towards the end of the calibrated period, which makes for slightly faster productivity growth. The saving and investment rates being virtually identical to the benchmark calibration at the mean and media, makes for a slightly increasing foreign surplus in line with Song et al.'s theory. The benchmark calibration by Song et al. can be backed up empirically when the estimates used are consistent with the model assumptions.

There is also one major similarity between Figure 2 and Figure 3: The alternative calibrations make for quite a lot of volatility. Volatility is the case for reallocation, saving, investment, the foreign reserve, and productivity growth, but we can especially note the investment rate and foreign reserve rate as quite varied across the alternative calibrations, even when the alternatives are consistent with the model assumptions. For the foreign reserve rate there are assumption-consistent parameter combinations that make for an increasing foreign reserve deficit, and there are calibrations for which the u-shape of the investment rate almost disappears. Too little reallocation makes for too little savings, which combined with too much capital investment makes for an increasing foreign reserve deficit. A low depreciation rate combined with a high capital share could make for such a dynamic.

Figure 3 – Assumption-consistent calibrations



Note: ‘alternatives’ refers to all the possible combinations of the alternative estimates of the capital share and depreciation rate that are consistent with the model assumptions, see Table 1, represented here by the black lines. The total number of black lines is thus 43. ‘mean alternatives’ is a calibration using the mean value of all consistent alternative calibrations, i.e. the mean of the black lines, here represented by the magenta line. Similarly, ‘median alternatives’ is a calibration using the median value of all consistent alternative calibrations, represented by the green line. The red line is the replicated benchmark calibration as presented in Figure 1.

Even if the model logic can explain the results in Figure 2 and 3, the degree of volatility in the foreign reserve rate subject to assumption-consistent alternative calibrations does not speak to the model being particularly suited to explain the Chinese foreign reserve rate. Small tweaks to the benchmark calibration can make for diametrically opposite predictions for the foreign reserve rate compared to the benchmark or extreme versions of the benchmark dynamic. The same holds for the investment rate and the rate of reallocation of labor, albeit not to the same degree. Figure 3 is, however, much more in line with, and speaks in favor of, the theory of Song

et al. and the calibration they do to illustrate their theory. The benchmark results of Song et al. are reasonably robust to alternative calibrations as long as the alternatives are consistent with the model assumptions.

The present thesis is limited in its scope in more than one way. There is undoubtedly relevant literature that I have failed to include on the subject of the capital share, depreciation rate and intertemporal elasticity of substitution as well as on calibration and growth due to misallocation. Consequently, the tests done on the basis of other literature are limited.

5.1.3 Model robustness subject to incremental changes of the benchmark values

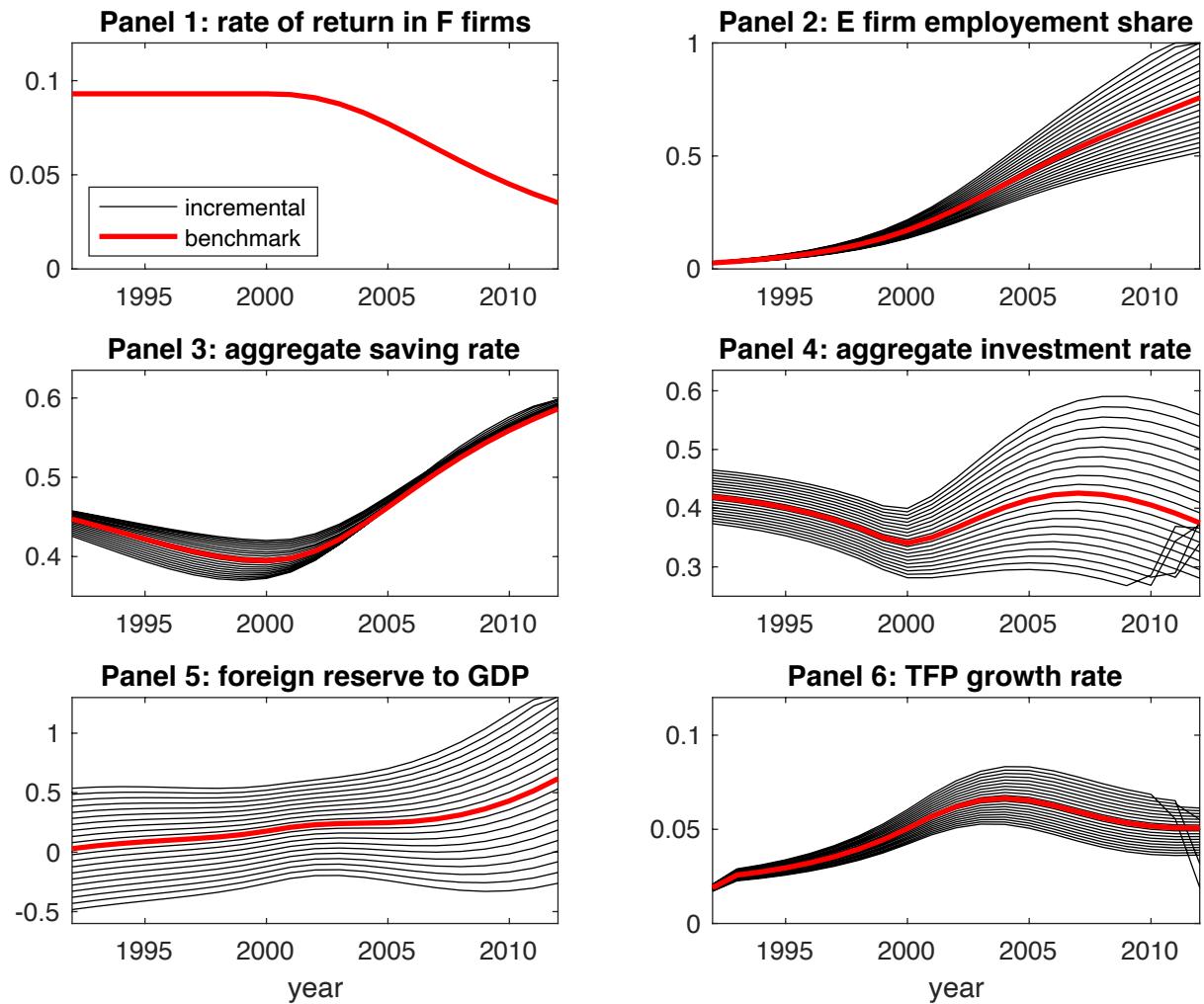
A final extension of the model asks the following question: How does the model react to small, isolated tweaks to the benchmark parameter values? The aim of such an incremental test is to test for model sensitivity. That is, to check whether the variation in the model is driven by specific parameters and whether the variation is inherent to the model, or a product of the estimates used for extensions. The tests work from the assumption of the original calibration is correct and illustrate the degree of uncertainty about the predictions of the model dependent on the capital share, depreciation rate, or intertemporal elasticity of substitution (Leamer, 1983; Simonsohn et al., 2020). The increments are calculated as one percent deviations from the benchmark parameter value for a twenty percent interval: Ten percent above the benchmark value and ten percent below the benchmark value. Each increment is calculated as:

$$\text{new value} = \text{benchmark} \pm (\text{benchmark} * \text{percentage})$$

The incremental tests are done separately for each parameter, making for three separate figures, each featuring twenty alternative calibrations and the benchmark calibration. The scope of the thesis makes it necessary to limit the size of the test; for three parameters and a twenty percent interval there would be 8000 possible alternative combinations.

Figure 4 shows the model results of an incremental test of the capital share. All variables and assumptions are recalculated for each incremental change. None of the twenty alternative calibrations are in violation of Assumptions 1-3.

Figure 4 – Incremental changes to the capital share benchmark value



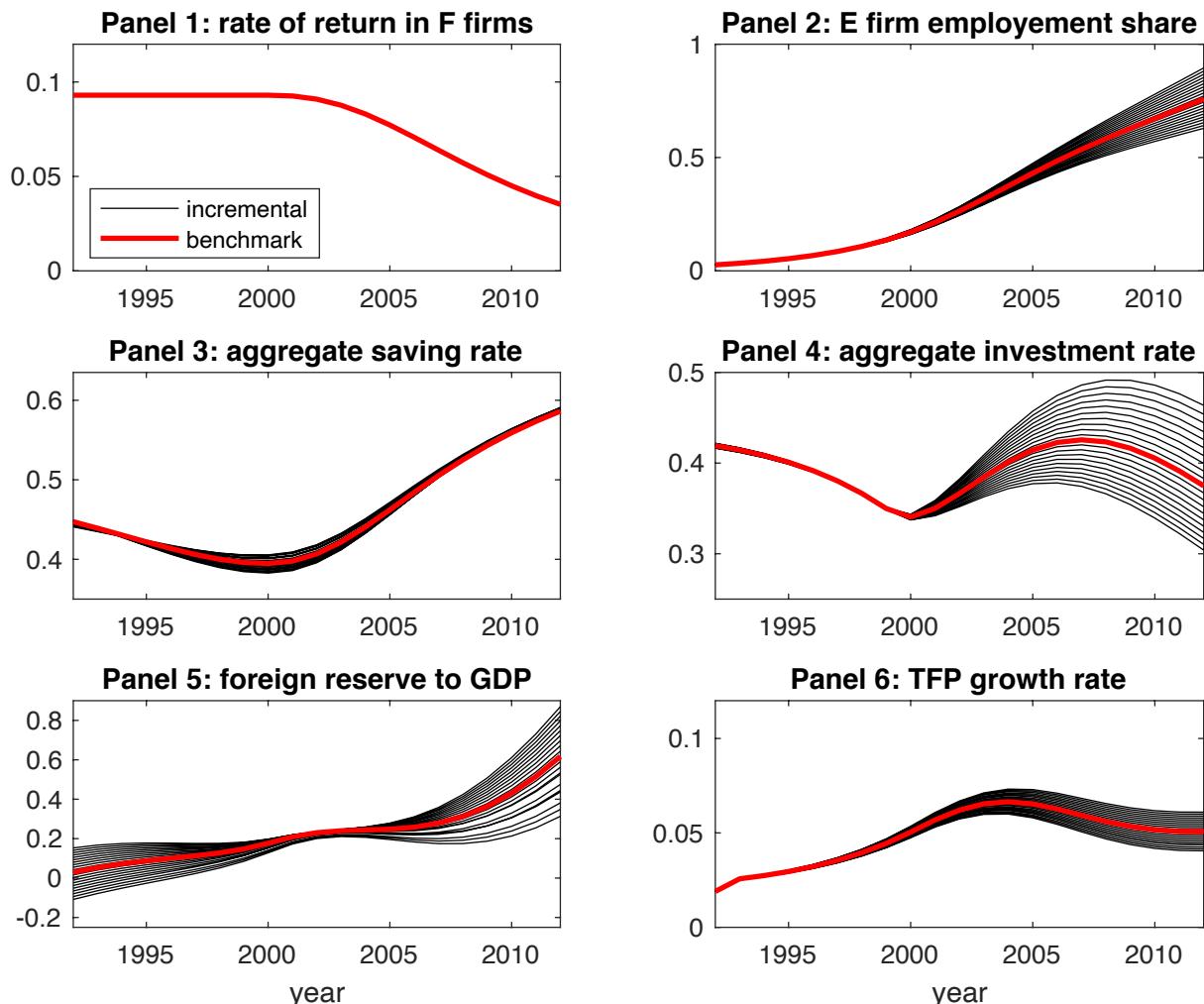
Note: The black lines represent incremental changes to the benchmark value by one percentage point over a \pm ten percent interval, hence ‘incremental’. ‘benchmark’, or the red line, represents the benchmark calibration as replicated in Figure 1 (Song et al., 2011, p. 223).

Figure 4 is illustrative of the influence of the capital share on the model dynamics. A higher capital share makes capital more important in production and so the reallocation of labor is slower and the investment rate higher. The saving rate does not change much for alternative capital shares, so the higher investment rate makes for a foreign reserve deficit. The deficit is somewhat decreasing due to financial development which leads to reallocation, more E-firms and more entrepreneurial saving. As the reallocation is slower for larger capital shares the productivity growth is also lower if very similar to the benchmark. The opposite of this interpretation is true for lower capital shares: Reallocation is faster, investments are lower, there is an increasing foreign reserve surplus and productivity growth is higher.

The test in Figure 4 underlines a feature of the tests in Figures 2 and 3: The investment and foreign reserve rates are quite sensitive to alternative calibration values. Sensitive relative to the other rates that is; the range of the investment and foreign reserve rate are by no means as large in Figure 4 as for the tests in Figures 2 and 3. As the increments are not large – the largest incremental change is ten percent, i.e., an upper-bound capital share of 0.55 and a lower-bound capital share of 0.45 – the volatility of the investment and foreign reserve rates is quite large.

Changing the depreciation rate of capital incrementally makes for less dramatic dynamics, as can be seen in Figure 5.

Figure 5 – Incremental changes to the depreciation rate benchmark value



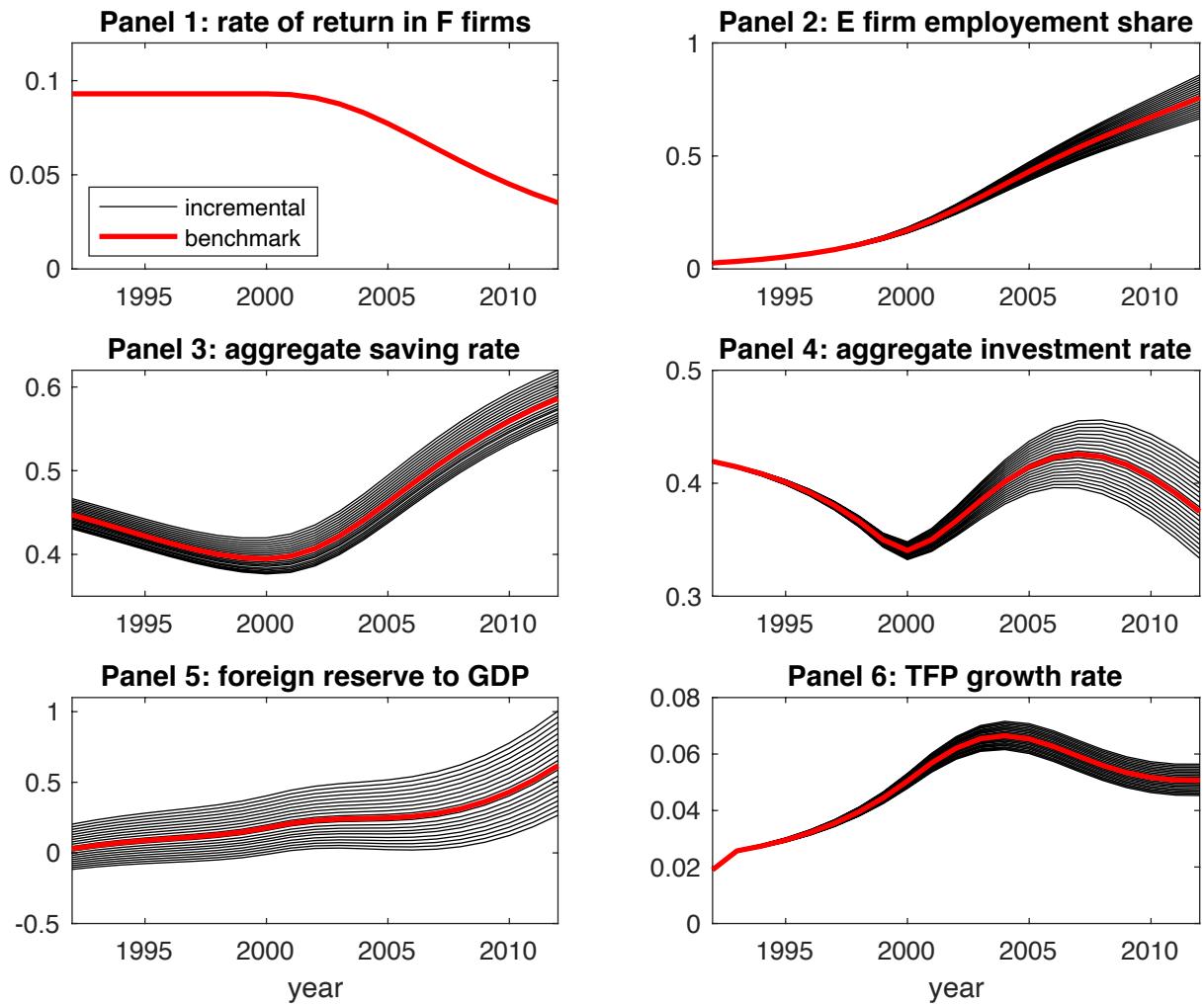
Note: The black lines represent incremental changes to the benchmark value by one percentage point over a ±ten percent interval, hence ‘incremental’. ‘benchmark’, or the red line, represents the benchmark calibration as replicated in Figure 1 (Song et al., 2011, p. 223).

The dynamics of the alternative calibrations in Figure 5 are largely identical to the benchmark. A higher capital depreciation rate makes for more investment to replace capital over time as well as a higher managerial share of income. The transition is therefore a little slower, as the larger managerial share of income makes for smaller returns to E-firms. The saving rate is more or less identical to the benchmark for all alternative calibrations, especially once there is financial development. Combined with investment rates not too different from the benchmark, the wanted dynamic of an increasing foreign reserve surplus is achieved but some of the larger depreciation rates make for an initial foreign reserve deficit. As the alternative reallocations of labor are very similar to the benchmark calibration, the alternative rates of productivity growth are very similar to the benchmark as well.

Figure 5 speaks to model robustness for changes to the capital depreciation rate alone. The changes are quite small, the largest deviation from the benchmark being one percentage point. The investment and foreign reserve rates are the most volatile series relative to the other rates in Figure 5, which by now seems to be true of any change to the benchmark calibration.

Figure 6 presents the results of a test using incremental changes to the benchmark intertemporal elasticity of substitution, $\theta = 2$. The upper-bound alternative calibration uses an elasticity of 2.2, while the lower-bound uses an elasticity of 1.8.

Figure 6 – Incremental changes to the benchmark intertemporal elasticity



Note: The black lines represent incremental changes to the benchmark value by one percentage point over a \pm ten percent interval, hence ‘incremental’. ‘benchmark’, or the red line, represents the benchmark calibration as replicated in Figure 1 (Song et al., 2011, p. 223).

The results in Figure 6 speak to reasonable model robustness when tweaking the intertemporal elasticity of substitution. The variation is larger than for tweaks to the depreciation rate, and somewhat smaller than tweaks to the capital share of income. First, we can note that the reallocation of labor is almost identical to the benchmark for all alternatives. Consequently, the productivity growth is almost identical for all alternatives as well. The positive percentage deviation from the benchmark makes for a saving rate that is a little higher, which makes for reallocation and productivity growth that are a little faster. The initial investment rate is identical to the benchmark for all alternatives, so a higher saving rate makes for a foreign surplus. As savings increase over the transition, investments do not increase as much since the need for capital does not increase much when the reallocation of labor is similar to the

benchmark. The foreign surplus is therefore growing as predicted by the model. The opposite logic is true for the negative percentage deviations: Lower saving makes for slower transition and productivity growth, while the initial investment level combined with lower initial saving makes for a foreign reserve deficit. As the saving rate increases with time, the foreign reserve deficit becomes an increasing surplus given a lower investment rate compared to the benchmark.

Again, there are some series that are more sensitive to incremental changes of the parameter values. In the case of the intertemporal elasticity of substitution, the more volatile series are the saving and foreign reserve rates. The effect of changing the elasticity is entirely within the model logic: A larger elasticity would make for more saving and a higher foreign reserve surplus by Assumption 3. Considering the magnitude of the changes however – 0.2 above and below the benchmark at the most – the foreign reserve rate seems quite volatile to small deviations, considering that the surplus is around one hundred percent at the end of the calibration period for the larger positive deviations. One concern is that the saving dynamic is too dependent on the value of the elasticity, making for incredible predictions for quite small tweaks to the initial value specification. Disregarding the volatility in Figure 6, the predictions of the Song et al.’s theory are achieved for all incremental changes to the benchmark elasticity value.

As with the tests I have done using alternative parameter estimates from relevant literature, the incremental test is limited in scope. It would have been relevant to calibrate the model for every possible combination of incremental changes. For the same interval of ten percent above and below the benchmark value, such a method would make for 8000 unique calibrations. One possibly interesting method for using a set of 8000 calibrations, would be to draw a random subset and test the model for it. Testing in the manner of random subsets would require more time than what I have had to work on my thesis. I therefore postpone the use of the complete incremental test to future research.

6 Conclusion

The general nature of calibration raises concerns about manipulation (Glaeser, 2006) and results driven by researcher choice (Leamer, 1983). The concerns about calibration do in part apply to the results of Song et al., as the choices of parameter values for the benchmark calibration are not discussed or justified by the authors in any major way.

That being said, the main results of the model in *Growing Like China* (Song et al., 2011) are somewhat robust to alternative calibrations. If one excludes all alternative parameter estimates that are inconsistent with the theoretical assumptions of the model, the median calibration of consistent alternatives is almost identical to the benchmark calibration. The mean calibration is qualitatively similar to the benchmark. There are assumption-consistent parameter estimates as well as combinations of alternative estimates that give results counter to the theory proposed by Song et al. (2011). It is a strength of the model that the benchmark results are quite robust to alternative calibrations as long as the alternatives are in line with the model assumptions. The model is however only robust if we accept the assumption that excludes the use of parameter values as estimated in the empirical literature (Atkeson & Ogaki, 1996; Attanasio & Weber, 1993, 1995; Campbell & Mankiw, 1989; Havránek, 2015; Havránek et al., 2015; Mankiw et al., 1985; Ogaki & Atkeson, 1997; Ogaki & Reinhart, 1998; Okubo, 2011).

My goal has not been to evaluate the theory of misallocation as an explanation of the Chinese growth experience. The theoretical arguments and the empirical evidence for reallocation are convincing (Chow, 1993; Hsieh & Klenow, 2009; Jones, 2015; Young, 2003). Given neoclassical theory it is also indeed puzzling that capital flows out of developing countries (Gourinchas & Jeanne, 2013) and the puzzle may well be due to factor reallocation and financial frictions (Song et al., 2011) at least in part (Gourinchas & Jeanne, 2013, pp. 1509–1510). The allocation puzzle is a feature of the model in the case of the mean and median value of assumption-consistent parameter estimates, and the median calibration is close to identical to the benchmark calibration. The level of volatility of the foreign reserve rate subjected to alternative calibrations is, however, very high.

The calibration in Song et al. and the robustness of the results may or may not be representative of other findings based in calibration. The findings here may therefore not be externally valid. My thesis is meant as an initiation of a much broader inquiry into the

robustness of calibration as a method. Once a broader set of models has been replicated and subjected to robustness tests in the manner outlined in my thesis, I will hopefully be able to make more concrete conclusions about the merits and demerits of calibration. I defer from making any sweeping conclusions about calibration as a method here, and merely state that the benchmark calibration in Song et al. can be tested quite extensively without the principal quantitative results being lost.

Literature

- Andersen, T. M. (1991). The Econometrics of the General Equilibrium Approach to Business Cycles: Comment. *The Scandinavian Journal of Economics*, 93(2), 179–184.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics : an empiricist's companion*. Princeton University Press.
- Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *The Journal of Economic Perspectives*, 24(2), 3–30. <https://doi.org/10.1257/jep.24.2.3>
- Atkeson, A., & Ogaki, M. (1996). Wealth-varying intertemporal elasticities of substitution: Evidence from panel and aggregate data. *Journal of Monetary Economics*, 38(3), 507–534. [https://doi.org/10.1016/S0304-3932\(96\)01290-1](https://doi.org/10.1016/S0304-3932(96)01290-1)
- Attanasio, O. P., & Weber, G. (1993). Consumption Growth, the Interest Rate and Aggregation. *The Review of Economic Studies*, 60(3), 631–649. <https://doi.org/10.2307/2298128>
- Attanasio, O. P., & Weber, G. (1995). Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey. *The Journal of Political Economy*, 103(6), 1121–1157. <https://doi.org/10.1086/601443>
- Bai, C.-E., Hsieh, C.-T., & Qian, Y. (2006). The Return to Capital in China. *Brookings Papers on Economic Activity*, 2006(2), 61–88. <https://doi.org/10.1353/eca.2007.0000>
- Beltran, D. O., & Draper, D. (2008). Estimating the Parameters of a Small Open Economy DSGE Model: Identifiability and Inferential Validity. *International Finance Discussion Papers*, 955, 40. <https://doi.org/10.2139/ssrn.1311887>
- Biederman, D. K., & Goenner, C. F. (2008). A life-cycle approach to the intertemporal elasticity of substitution. *Journal of Macroeconomics*, 30(1), 481–498. <https://doi.org/10.1016/j.jmacro.2006.05.002>
- Blanchard, O. (2016). Do DSGE models have a future? *Revista de Economia Institucional*, 18(35), 39–46. <https://doi.org/10.18601/01245996.v18n35.03>
- Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics*, 8(1), 1–32. <https://doi.org/10.1257/app.20150044>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability

- of laboratory experiments in economics. *Science*, 351(6280), 1433 LP – 1436.
<https://doi.org/10.1126/science.aaf0918>
- Campbell, J. Y., & Mankiw, N. G. (1989). Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence. *NBER Macroeconomics Annual*, 4, 185–216.
<https://doi.org/10.1086/654107>
- Canova, F., & Sala, L. (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4), 431–449.
<https://doi.org/10.1016/j.jmoneco.2009.03.014>
- Chow, G. C. (1993). Capital Formation and Economic Growth in China. *The Quarterly Journal of Economics*, 108(3), 809–842. <https://doi.org/10.2307/2118409>
- DeJong, D. N. (2011). *Structural macroeconomics* (C. Dave (ed.); 2nd ed.). Princeton University Press.
- Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *The American Economic Review*, 76(4), 587–603. <http://www.jstor.org/stable/1806061>
- Diao, X., McMillan, M., & Rodrik, D. (2017). *The Recent Growth Boom in Developing Economies: A Structural Change Perspective*.
- Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What Is Meant by “Replication” and Why Does It Encounter Resistance in Economics? *American Economic Review*, 107(5), 46–51. <https://doi.org/10.1257/aer.p20171031>
- Finansdepartementet. (2019). *Rapport om utviklingen av NORA – en ny makroøkonomisk modell for analyse av finanspolitikken*.
https://www.regjeringen.no/globalassets/upload/fin/oa/nora_rapport_utviklingen.pdf
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502 LP – 1505.
<https://doi.org/10.1126/science.1255484>
- Frisch, R. (1933a). *Propagation problems and impulse problems in dynamic economics: Vol. nr 3*. Universitetets økonomiske institutt.
- Frisch, R. (1933b). Editor’s Note. *Econometrica*, 1(1), 1–4.
<http://www.jstor.org/stable/1912224>
- Glaeser, E. L. (2006). *Researcher Incentives and Empirical Methods* (NBER Technical Working Paper Series). National Bureau of Economic Research.
<https://doi.org/10.3386/t0329>
- Gourinchas, P.-O., & Jeanne, O. (2013). Capital Flows to Developing Countries: The

- Allocation Puzzle. *The Review of Economic Studies*, 80(4 (285)), 1484–1515.
<https://doi.org/10.1093/restud/rdt004>
- Guerriero, M. (2019). The labor share of income around the world: Evidence from a panel dataset. In *ADBI Working Paper Series* (No. 920). <https://doi.org/10.2139/ssrn.3358006>
- Hall, R. E. (1988). Intertemporal Substitution in Consumption. *The Journal of Political Economy*, 96(2), 339–357. <https://doi.org/10.1086/261539>
- Hansen, L. P., & Heckman, J. J. (1996). The Empirical Foundations of Calibration. *The Journal of Economic Perspectives*, 10(1), 87–104. <https://www.jstor.org/stable/2138285>
- Havránek, T. (2015). Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting. *Journal of the European Economic Association*, 13(6), 1180–1204. <https://doi.org/10.1111/jeea.12133>
- Havránek, T., Horvath, R., Irsova, Z., & Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, 96(1), 100–118. <https://doi.org/10.1016/j.inteco.2015.01.012>
- He, Q., Liu, F., Qian, Z., & Tai Leung Chong, T. (2017). Housing prices and business cycle in China: A DSGE analysis. *International Review of Economics & Finance*, 52, 246–256. <https://doi.org/10.1016/j.iref.2017.01.012>
- Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4), 1403–1448. <https://doi.org/10.1162/qjec.2009.124.4.1403>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Med*, 2(8), 696–701. <https://doi.org/10.1371/journal.pmed.0020124>
- Jones, C. I. (2015). *The Facts of Economic Growth* (No. 21142; NBER Working Papers). <http://www.nber.org/papers/w21142>
- Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *American Economic Review*, 106(9), 2426–2457. <https://doi.org/10.1257/aer.20110236>
- Kydland, F. E., & Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica*, 50(6), 1345–1370. <https://doi.org/10.2307/1913386>
- Kydland, F. E., & Prescott, E. C. (1991). The Econometrics of the General Equilibrium Approach to Business Cycles. *The Scandinavian Journal of Economics*, 93(2), 161–178. <https://doi.org/10.2307/3440324>
- Kydland, F. E., & Prescott, E. C. (1996). The Computational Experiment: An Econometric Tool. *The Journal of Economic Perspectives*, 10(1), 69–85. <https://doi.org/10.1257/jep.10.1.69>

- Leamer, E. E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review*, 73(1), 31–43.
- Leamer, E. E. (2010). Tantalus on the Road to Asymptopia. *The Journal of Economic Perspectives*, 24(2), 31–46. <https://doi.org/10.1257/jep.24.2.31>
- Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(C), 19–46. [https://doi.org/10.1016/S0167-2231\(76\)80003-6](https://doi.org/10.1016/S0167-2231(76)80003-6)
- Lucas, R. E. (1980). Methods and Problems in Business Cycle Theory. *Journal of Money, Credit and Banking*, 12(4), 696–715. <https://doi.org/10.2307/1992030>
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2), 407–437. <https://doi.org/10.2307/2118477>
- Mankiw, N. G., Rotemberg, J. J., & Summers, L. H. (1985). Intertemporal Substitution in Macroeconomics. *The Quarterly Journal of Economics*, 100(1), 225–251. <https://doi.org/10.2307/1885743>
- McCullough, B. D., McGahey, K. A., & Harrison, T. D. (2006). Lessons from the JMBCB Archive. *Journal of Money, Credit and Banking*, 38(4), 1093–1107. <http://www.jstor.org/stable/3838995>
- OECD. (2000). *National Accounts for China - Sources and Methods* (E. Burgeat (ed.)). Organisation for Economic Co-operation and Development.
- Ogaki, M., & Atkeson, A. (1997). Rate of Time Preference, Intertemporal Elasticity of Substitution, and Level of Wealth. *The Review of Economics and Statistics*, 79(4), 564–572. <https://doi.org/10.1162/003465397557141>
- Ogaki, M., & Reinhart, C. M. (1998). Measuring Intertemporal Substitution: The Role of Durable Goods. *The Journal of Political Economy*, 106(5), 1078–1098. <https://doi.org/10.1086/250040>
- Okubo, M. (2011). The Intertemporal Elasticity of Substitution: An Analysis Based on Japanese Data. *Economica (London)*, 78(310), 367–390. <https://doi.org/10.1111/j.1468-0335.2009.00826.x>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263(5147), 641–646. <https://doi.org/10.1126/science.263.5147.641>

- Perkins, D. H. (1988). Reforming China's Economic System. *Journal of Economic Literature*, 26(2), 601–645.
- Podpiera, R. (2006). *Progress in China's banking sector reform : has bank behavior changed?* (WP/06/71; IMF Working Paper (Online)).
<https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Progress-in-Chinas-Banking-Sector-Reform-Has-Bank-Behavior-Changed-18952>
- Prescott, E. C. (1986). Theory ahead of business-cycle measurement. *Carnegie-Rochester Conference Series on Public Policy*, 25(C), 11–44. [https://doi.org/10.1016/0167-2231\(86\)90035-7](https://doi.org/10.1016/0167-2231(86)90035-7)
- Romer, P. (2016). The Trouble With Macroeconomics. *Commons Memorial Lecture of the Omicron Delta Epsilon Society*, 26. <https://doi.org/10.1192/pb.33.10.396a>
- Schündeln, M. (2013). Appreciating depreciation: physical capital depreciation in a developing country. *Empirical Economics*, 44(3), 1277–1290.
<https://doi.org/10.1007/s00181-012-0592-2>
- Shoven, J. B., & Whalley, J. (1992). *Applying General Equilibrium*. Cambridge University Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychol Sci*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simon, J. L., & Aigner, D. J. (1970). Cross-Section and Time-Series Tests of the Permanent-Income Hypothesis. *The American Economic Review*, 60(3), 341–351.
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-020-0912-z>
- Sims, C. A. (1996). Macroeconomics and Methodology. *The Journal of Economic Perspectives*, 10(1), 105–120. <https://doi.org/10.1257/jep.10.1.105>
- Smets, F., & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *The American Economic Review*, 97(3), 586–606.
<https://doi.org/10.1257/aer.97.3.586>
- Song, Z., Storesletten, K., & Zilibotti, F. (2011). Growing like China. *American Economic Review*, 101(1), 196–233. <https://doi.org/10.1257/aer.101.1.196>
- Tobin, J. (1950). A Statistical Demand Function for Food in the U.S.A. *Journal of the Royal Statistical Society. Series A. General*, 113(2), 113–149. <https://doi.org/10.2307/2981038>
- Wang, Y., & Yao, Y. (2003). Sources of China's economic growth 1952–1999: incorporating human capital accumulation. *China Economic Review*, 14(1), 32–52.

[https://doi.org/10.1016/S1043-951X\(02\)00084-6](https://doi.org/10.1016/S1043-951X(02)00084-6)

Young, A. (2003). Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period. *The Journal of Political Economy*, 111(6), 1220–1261.
<https://doi.org/10.1086/378532>

Appendix

A - Growing Like China presented analytically

Life-time utility of consumption is given on constant relative risk aversion (CRRA) form:

$$U_t = \frac{(c_{1t})^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}} + \beta \frac{(c_{2t+1})^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}} \quad (1)$$

The labor force is assumed to equal the general population and grows at an exogenously given rate:

$$N_{t+1} = (1 + \nu)N_t$$

We treat N_t as workers, while a fraction of each cohort, μN_t , are entrepreneurs. Firm-level production is given as Harrod-neutral Cobb-Douglas functions:

$$y_{Ft} = k_{Ft}^\alpha (A_t n_{Ft})^{1-\alpha}$$

$$y_{Et} = k_{Et}^\alpha (\chi A_t n_{Et})^{1-\alpha}$$

χ is the productivity advantage of E-firms and managerial input is accounted for by n_{Et} . General productivity grows at an exogenously given rate:

$$A_{t+1} = (1 + z)A_t$$

Going back to the life-cycle utility function in (1), we assume that workers save some share of their earnings, w , at a gross interest rate, R^d . Optimal savings are found by maximizing (1) subject to a standard intertemporal budget constraint, $\frac{c_{1t}^W + c_{2t+1}^W}{R^d} = w_t$. Through insertion we then obtain:

$$s_t^W = \zeta^W w_t, \quad \zeta^W \equiv (1 + \beta^{-\theta} R^{1-\theta})^{-1}$$

The young entrepreneurs earn a managerial compensation to be defined below, m , which can either be saved in a bank account or reinvested in the firm they work for.

The financial sector collects savings and invests either in the form of loans to domestic firms or in foreign bonds. It is assumed that the output of E-firms is non-verifiable and that entrepreneurs will only be able to pledge to repay a share, η , of net profits in the second period. Operational costs, red tape etc. amounts to an iceberg cost of lending to domestic firms, ξ . Finally, the gross return on foreign bonds is R . In a competitive equilibrium the rate of return on domestic loans, R^d , must equal the rate of return on foreign bonds. All in all, we have equilibrium the rate of return the lending rate:

$$R^d = R \wedge R^l = \frac{R}{1 - \xi}$$

Profit maximization for F-firms implies that the marginal product of capital is equal to the lending rate to domestic firms, R^l , and that the marginal product of labor equals the wage rate:

$$\frac{\partial y_{Ft}}{\partial k_{Ft}} = R^l \Leftrightarrow R^l = \alpha \left(\frac{k_{Ft}}{A_t n_{Ft}} \right)^{\alpha-1} \leftrightarrow \frac{k_{Ft}}{A_t n_{Ft}} = \left(\frac{R^l}{\alpha} \right)^{\frac{1}{1-\alpha}}$$

$$\frac{\partial y_{Ft}}{\partial n_{Ft}} = w_t \Leftrightarrow w_t = (1 - \alpha) \left(\frac{k_{Ft}}{A_t n_{Ft}} \right)^\alpha A_t$$

$$w_t = (1 - \alpha) \left(\frac{R^l}{\alpha} \right)^{\frac{\alpha}{1-\alpha}} A_t \quad (2)$$

With (2) we can determine the value of an E-firm. Recall that the managerial compensation will be a fraction, ψ , of output. An E-firm owner will then have the following optimization problem for a given level of capital:

$$\Xi_t(k_{Et}) = \max_{n_{Et}} \{y_{Et} - m_t - w_t n_{Et}\} \mid m_t = \psi y_{Et}, \quad y_{Et} = k_{Et}^\alpha (\chi A_t n_{Et})^{1-\alpha} \quad (3)$$

The first-order condition of (3) yields the optimal E-firm employment at the equilibrium wage given by (2):

$$n_{Et} = [(1 - \psi)\chi]^{\frac{1}{\alpha}} \left(\frac{k_{Et}}{\chi A_t} \right) \left(\frac{R^l}{a} \right)^{\frac{1}{1-\alpha}} \quad (4)$$

The optimal value of an E-firm is found by plugging (4) into (3) and using our definition of the managerial compensation:

$$\Xi_t(k_{Et}) = (1 - \psi)^{\frac{1}{\alpha}} \chi^{\frac{1-\alpha}{\alpha}} R^l k_{Et} \equiv \rho_E k_{Et} \quad (5)$$

If the model is to yield a reallocation from F- to E-firms the rate of return in E-firms must be larger than the lending rate to domestic firms, $\rho_E > R^l$. Hence, the following assumption:

$$\text{ASSUMPTION 1: } \chi > \underline{\chi} \equiv \left(\frac{1}{1-\psi} \right)^{\frac{1}{1-\alpha}} \quad (6)$$

The capital stock of an E-firm is given by the savings of the young entrepreneurs and the firm's loan, $k_{Et} = s_{Et-1} + l_{Et-1}$. Given (6), (5) and that E-firms can only guarantee to repay a share of income for any loan, we have:

$$R^l l_E \leq \eta \rho_E k_E \Leftrightarrow l_E (R^l - \eta \rho_E) \leq \eta \rho_E s_E$$

The constraint that ensures entrepreneurs' reinvestment in E-firms is thus only binding if $\eta < \frac{R^l}{\rho_E}$, which is assumed henceforth. The share of investment financed by bank loans is then given by:

$$\frac{l_E}{l_E + s_E} = \frac{\eta \rho_E}{R^l} \Leftrightarrow l_E = \frac{\eta \rho_E}{R^l - \eta \rho_E} s_E \quad (7)$$

The optimal saving rate can now be found from the intertemporal utility of consumption as defined in (1), using (7) and the fact that entrepreneurs' consumption must be $c_1 = m - s_E$ and $c_2 = \rho_E k_E - R^l l_E$ (dropping the time scripts for now):

$$\begin{aligned}
\max_{s_E} U &= \max_{s_E} \frac{(m - s_E)^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}} + \beta \frac{(\rho_E k_E - R^l l_E)^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}} \\
&\Leftrightarrow \max_{s_E} \frac{(m - s_E)^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}} + \beta \frac{\left(\frac{(1-\eta)R^l\rho_E}{R^l - \eta\rho_E}s_E\right)^{1-\frac{1}{\theta}} - 1}{1 - \frac{1}{\theta}}
\end{aligned}$$

The problem implies:

$$s_E = \left\{ 1 - \beta^{-\theta} \left[\frac{(1-\eta)R^l\rho_E}{R^l - \eta\rho_E} \right]^{1-\theta} \right\} m \equiv \zeta^E m$$

We now move on to describing the process of transition and the possible equilibriums reached dependent on the degree of saving by entrepreneurs. Due to the difficulties E-firms have in obtaining external financing, there will be a lower ratio of capital to output in E-firms. Let efficient capital be denoted by κ . Then profit maximization in F-firms as found above is rewritten:

$$\kappa_F \equiv \frac{k_F}{A_F n_F} = \left(\frac{\alpha}{R^l} \right)^{\frac{1}{1-\alpha}} \quad (8)$$

By (8) there is a constant optimal level of capital. Wages grow at the rate of productivity growth, z , as given by (2). The level of efficient capital in E-firms is found by using (8) in the definition of labor employed in E, given by (4):

$$\begin{aligned}
n_E &= [(1 - \psi)\chi]^{\frac{1}{\alpha}} \left(\frac{k_E}{\chi A} \right) \left(\frac{R^l}{a} \right)^{\frac{1}{1-\alpha}} \\
\left(\frac{k_E}{\chi A n_E} \right) &= [(1 - \psi)\chi]^{-\frac{1}{\alpha}} \left(\frac{\alpha}{R^l} \right)^{\frac{1}{1-\alpha}} \\
\kappa_E &= \kappa_F [(1 - \psi)\chi]^{-\frac{1}{\alpha}} \quad (9)
\end{aligned}$$

LEMMA 1: *Let assumption 1 hold. Then E-firms will have a lower capital-output ratio and capital-labor ratio than F-firms.*

Before we consider the equilibrium, we must note three key properties of the model that makes for constant growth of employment, capital and output in E-firms during transition:

- E-firm capital and productivity are state variable, while F-firm capital follows from profit maximization.
- The level of capital per efficient unit of labor is constant for both types of firms.
- The savings of entrepreneurs in period any period is linear in the current level of capital.

From these properties we are able to state the equilibrium dynamics of the model as¹²:

LEMMA 2: *Given K_{Et} and A_t , the equilibrium dynamics of total capital and employment in E-firms during transition are given as:*

$$\frac{K_{Et+1}}{K_{Et}} = (1 + \gamma_{K_E}) \wedge \frac{N_{Et+1}}{N_{Et}} = \frac{(1 + \gamma_{K_E})}{(1 + z)} \equiv 1 + v_E$$

where

$$1 + \gamma_{K_E} = \frac{R^l}{R^l - \eta\rho_E} \left[1 + \beta^{-\theta} \left(\frac{(1 - \eta)\rho_E R^l}{R^l - \eta\rho_E} \right)^{1-\theta} \right]^{-1} \frac{\psi}{1 - \psi} \frac{\rho_E}{\alpha} \quad (10)$$

and

$$\rho_E = R^l (1 - \psi)^{\frac{1}{\alpha}} \chi^{\frac{1-\alpha}{\alpha}}$$

There exists a critical value for the efficiency advantage of E-firms that must be exceeded for the employment share of E-firms, $\frac{N_E}{N}$, to grow over time. Assuming higher employment growth in E-firms and using (10) we get:

$$v_E > v \Leftrightarrow 1 + \gamma_{K_E} > (1 + v)(1 + z)$$

¹² Due to the constant returns to scale as implied by the production functions, we merely scale up from firm to aggregate level in this part.

$$\begin{aligned}
& \frac{\psi}{1 - \psi} \frac{1}{\alpha(1 + v)(1 + z)} \\
& > \frac{1}{R^l} \left[(1 - \psi)^{-\frac{1}{\alpha}} \chi^{-\frac{1-\alpha}{\alpha}} - \eta \right] \\
& + \beta^{-\theta} (1 - \eta)^{1-\theta} \left\{ \frac{1}{R^l} \left[(1 - \psi)^{-\frac{1}{\alpha}} \chi^{-\frac{1-\alpha}{\alpha}} - \eta \right] \right\}^\theta \quad (11)
\end{aligned}$$

The efficiency advantage that gives equality in (11) is the critical value which the actual efficiency gains must exceed for there to be an increasing share of employment in E-firms:

$$\chi > \hat{\chi} = \hat{\chi}(\beta, \chi, \psi, \eta, \alpha, v, z, R, \xi) < \infty$$

The critical value is decreasing in β , η and R^l and increasing in v and z . For there to be a transition then, the discount factor on consumption, the ability of entrepreneurs to repay loans and the lending rate must be sufficiently large, and the population and productivity growth sufficiently small for there to be growing employment in E-firms.

We have two further assumptions that must be noted. Capital is implied to grow at a constant rate here, which hinges on:

- Constant return to capital in every E firm.
- The earnings and savings of young entrepreneurs are proportional to the profit of E-firms.

If these assumptions hold, there will be growing income inequality between workers and entrepreneurs. Suppose zero productivity growth. Then wages will not grow while managerial compensation grows at the rate of E-firm investment:

$$\begin{aligned}
\frac{w_{t+1}}{w_t} &= \frac{A_{t+1}}{A_t} = (1 + z) = 1 \\
\frac{m_{t+1}}{m_t} &= \left(\frac{K_{Et+1}}{K_{Et}} \right)^\alpha \left(\frac{A_{t+1}}{A_t} \right)^{1-\alpha} \left(\frac{N_{Et+1}}{N_{Et}} \right)^{1-\alpha} = 1 + \gamma_{K_E}
\end{aligned}$$

Growing inequality is key to the transition as investment in E-firms is financed by entrepreneurial savings and constant wages ensures that the returns on investment are not

falling. The role of the managerial share of income, ψ , can be shown by using the definition of E-firm returns, ρ_E , into (10):

$$1 + \gamma_{K_E} = \frac{1}{1 - \eta} \frac{1}{\alpha} \frac{\psi}{1 - \psi} \frac{1}{R^l + (\beta R^l)^{-\theta}}$$

Hence, the growth of E-firms is hump-shaped in the managerial share of income: A smaller share makes for poorer entrepreneurs and less investment, while a sufficiently large income share leads falling profitability in E-firms. There is a sweet-spot where E-firms thrive.

Equilibrium dynamics for F-firms can be found residually from the condition that F-firms hire any workers not employed by E-firms and that capital adjusts optimally according to the capital-labor ratio: $K_{Ft} = \kappa_F A_t (N_t - N_{Et})$. From this alone it should be clear that an increasing employment share of E-firms leads to decreasing capital growth in F-firms; capital growth is hump-shaped in N_{Et} .

GDP per worker can be found using (8) and (9):

$$\frac{Y_t}{N_t} = \frac{Y_{Ft} + Y_{Et}}{N_t} = \frac{\kappa_F^\alpha A_t N_{Ft} + \kappa_F^\alpha A_t N_{Et} (1 - \psi)^{-1}}{N_t} = \kappa_F^\alpha \left(1 - \frac{\psi}{1 - \psi} \frac{N_{Et}}{N_t} \right) A_t \quad (12)$$

(12) is growing all throughout the transition due to reallocation of inputs towards the more efficient E-firms, i.e. as long as $\chi > \hat{\chi}$.

The banks' balance sheet is given the standard way: Assets must equal liabilities. Recall that the fraction of loans to E-firms is given by (7). Then we have:

$$B_t + K_{Ft} + \frac{\eta \rho_E}{R^l} K_{Et} = \zeta^W w_{t-1} N_{t-1} \quad (13)$$

Given the definitions of capital accumulation, wage growth and population growth presented thus far, there is a third lemma:

LEMMA 3: *The foreign surplus is given by*

$$B_t = \left(\zeta^W \frac{(1-\alpha)\kappa_F^{\alpha-1}}{(1+z)(1+v)} - 1 + (1-\eta) \frac{N_{Et}}{N_t} \right) \kappa_F A_t N_t \quad (14)$$

(14) implies a growing foreign surplus measured in efficiency units during a reallocation of workers to E-firms. If E-firms are sufficiently credit constrained, the foreign surplus will also be positive once the transition has ended. As less and less savings are needed to finance domestic investment, capital migrates abroad. That necessarily means that the credit constraint must be sufficiently severe, and the managerial share of income must be sufficiently small:

$$\frac{\psi}{1-\eta(1-\psi)} < \frac{\alpha(1+z)(1+v)}{(1-\alpha)} \frac{1+\beta^{-\theta} R^{1-\theta}}{R^l}$$

The reason the reallocation of the labor force works on the foreign surplus in this way is that the saving rate and investment rate, given the assumptions, are increasing and falling respectively during the transition. The growth of aggregate savings is due to:

- Workers employed by F-firms earn $(1-\alpha)$ and save ζ^W while the workers employed in E-firms save $\zeta^W(1-\alpha)(1-\psi)$.
- The entrepreneurs save $\zeta^E\psi$.

All of this makes for savings-output ratio in E-firms of:

$$\zeta^W(1-\alpha) + \alpha\zeta^E\psi + (1-\alpha)\psi(\zeta^E - \zeta^W)$$

which will be larger than the saving rate in F-firms as long as $\zeta^E \geq \zeta^W$. The saving rate of entrepreneurs is only larger than the saving rate of workers if $\theta \geq 1$, as the return in E-firms being larger than the economy lending rate leads to entrepreneurs' savings increasing more than workers' savings.

As for the decreasing investment rate, the intuition is that, assuming no population or productivity growth, any worker moved from F to E works with less capital. Domestic investments thus fall, but this is not necessary for the foreign surplus to increase during the transition.

At the end of the transitional period all workers are employed in E-firms and the model collapses to a standard OLG case. Consider log-preferences, $\theta \rightarrow 1$. The aggregate capital stock follows from the capital accumulation in E-firms:

$$k_{Et+1} = s_{Et} + l_{Et} \Leftrightarrow K_{Et+1} = \left(\frac{\beta}{1-\beta} \right) \left(\frac{R^l}{R^l - \eta\rho_E} \right) m_t$$

The post-transition law of motion is obtained by inserting the definition of m_t given by (3) and noting that the rate of return in E-firms now follows directly from the marginal product of capital:

$$Y_{Et} = (1-\psi)K_{Et}^\alpha (\chi A_t N_{Et})^{1-\alpha} \rightarrow \rho_E = \frac{dY_{Et}}{dK_{Et}} = \alpha(1-\psi)\kappa_{Et}^{\alpha-1}$$

$$\kappa_{Et+1} = \left(\frac{\beta}{1-\beta} \right) \left(\frac{\psi}{(1+z)(1+\nu)} \right) \left(\frac{R^l}{R^l - \eta\alpha(1-\psi)\kappa_{Et}^{\alpha-1}} \right) \kappa_{Et}^\alpha \quad (15)$$

By (15) capital deepens until the rate of return equals the lending rate to domestic firms or capital in efficiency units converges to a steady state with capital returns exceeding said lending rate. During convergence efficiency unit wages and output, and the foreign surplus are increasing as the rate of return to capital falls.

Allowing for a decreasing iceberg cost of financial intermediation makes for a decreasing lending rate over time:

$$R^l = \frac{R}{1-\xi_t}$$

Ceteris paribus, the capital-labor ratio and the wages in both types of firms will increase during the transition.

B – Example of code for incremental test

I have chosen to include the code for one alternative calibration using incremental changes to the intertemporal elasticity of substitution. The complete script for the test is included in the separate appendix as *robustness_increments_sigma*.

```
+1%
clear all

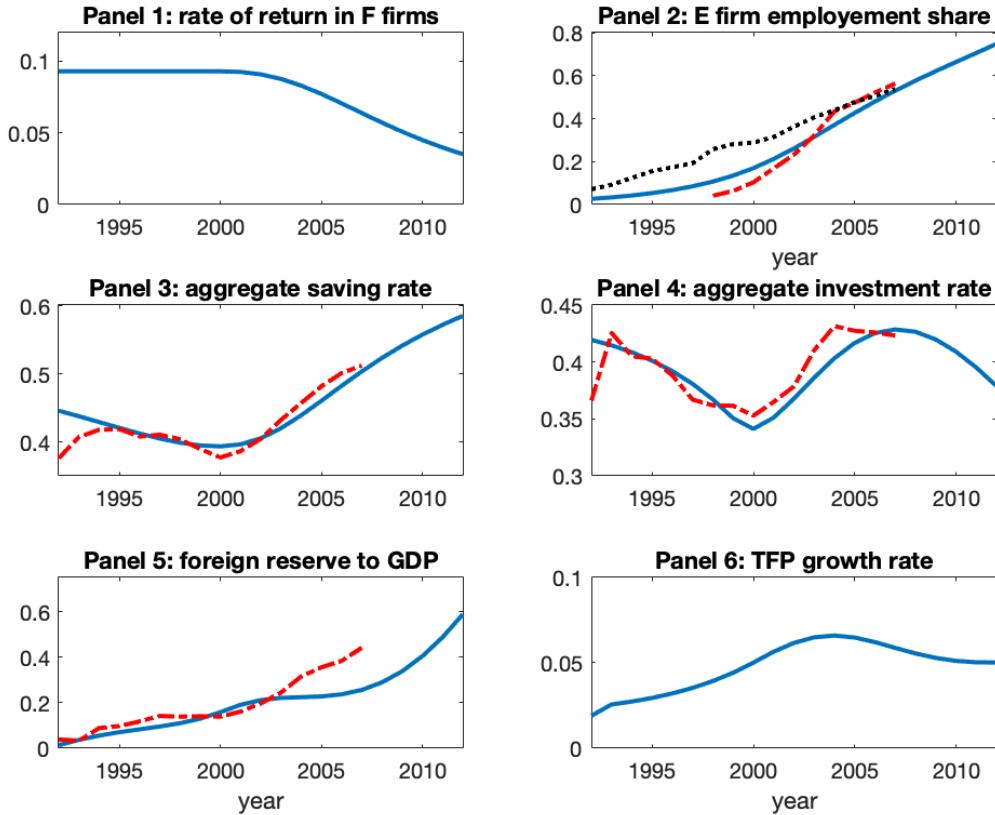
run parameter
Assumption 1 > 0
ans = 1.5231
sig = sig + (sig*0.01); %Constant relative risk aversion

bet = 0.998;
bet_E = bet;
bet = bet*(1+g_t)^(1-sig); %Discount factor(s)
bet_E = bet_E*(1+g_t)^(1-sig);

psi = 1 - (rho_r*r/(1-ice) + del)/(r/(1-ice) + del)/KY_F_E %Managerial
compensation
psi = 0.4467
ksi = (KY_F_E)^(alp/(1-alp))/(1-psi) %Productivity advantage
ksi = 4.7892
disp('Assumption 1 > 0')
Assumption 1 > 0
ksi = (1/(1-psi))^(1/(1-alp)) %Assumption 1
ans = 1.5231
disp('Assumption 2 > 0')
Assumption 2 > 0
(psi/(1-psi))*1/(alp*(1+g_n)*(1+g_t)) - ((1-ice)/r)*(((1-psi)^(-
1/alp))*ksi^(-(1-alp)/alp) - eta) - bet^(-1/sig)*(1-eta)^(1-1/sig)*((1-
ice)/r)*(((1-psi)^(-1/alp))*ksi^(-(1-alp)/alp) - eta)^(1/sig)
%Assumption 2
ans = 1.1276 + 0.1421i
disp('Assumption 3 > 0')
Assumption 3 > 0
((alp*(1+g_n)*(1+g_t))/(1-alp))*(1+bet^(-1/sig)*r^(1-1/sig))/(r/(1-
ice)) - psi/(1-eta*(1-psi)) %Assumption 3
ans = 597.5496
run pre_transition

run transition
```

```
run six_panel
Warning: Imaginary parts of complex X and/or Y arguments ignored.
Warning: Imaginary parts of complex X and/or Y arguments ignored.
Warning: Imaginary parts of complex X and/or Y arguments ignored.
Warning: Imaginary parts of complex X and/or Y arguments ignored.
Warning: Imaginary parts of complex X and/or Y arguments ignored.
```



```
r_F_incsig1 = r_F;
NE_N_t_incsig1 = NE_N_t;
S_Y_t_incsig1 = S_Y_t;
I_Y_t_incsig1 = I_Y_t;
FA_Y_t_incsig1 = FA_Y_t;
TFP_t_incsig1 = TFP_t;

save data_incsig1 r_F_incsig1 NE_N_t_incsig1 S_Y_t_incsig1
I_Y_t_incsig1 FA_Y_t_incsig1 TFP_t_incsig1
```