



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Essays on climate change, energy, and independence

David Comerford

DOCTOR OF PHILOSOPHY

in Economics

UNIVERSITY OF EDINBURGH

2013

Abstract

This thesis contains three separate papers. *A balance of questions: what can we ask of climate change economics?* is a critical analysis of the economics of climate change literature. It concludes that much more research effort needs to be put into studying the investment needed for a transition to a zero carbon energy infrastructure, rather than the focus on determining the social cost of carbon. *The interaction of scale economies and energy quality* is a theoretical study of the ability of economies to operate given different qualities of energy resources. *Measuring costs and benefits of independence* is an analysis of the welfare costs to Catalonia from reduced trade, which may arise on independence from Spain. These costs are set against the benefits to Catalonia of not paying fiscal transfers to the rest of Spain.

Acknowledgements

I would like to thank my supervisor, Dr Richard Holt, and co-author Prof Sevi Mora for the great deal of help and encouragement they have given me throughout the PhD process. I would like to thank my external examiner Prof John Hassler, and internal examiner Dr Robert Zymek for some great comments and suggestions during the viva, and I would also like to thank Prof Stuart Sayer, Prof Maia Guell, Dr Sean Brocklebank, Prof Ed Hopkins, Prof Andy Snell, Dr Simon Clark, Dr Ahmed Anwar, Prof Michele Belot, Dr Laurence Lasselle, Dr Ian Lange, Mrs Lorna Aitken, Miss Eirlys Armstrong, Ms Marie Craft, Mrs Lesley Mayne, Ms Gina Reddie, Mr Nick Myers, Mr David Pugh, Mr James Best, Dr Jelte Harnmeijer & Mr Andre Loureiro.

Contents

Chapter	Page No.
- Declaration	3
1. A balance of questions: what can we ask of climate change economics?	4
- Abstract	4
- Article	5
- References	25
- Appendices	27
2. The interaction of scale economies and energy quality	35
- Abstract	35
- Article	36
- References	66
- Appendices	68
3. Measuring Costs and Benefits of Independence	77
- Abstract	77
- Article	78
- References	104
- Appendices	105

Declaration

This is to certify that that the work contained within has been composed by me and is entirely my own work apart from the chapter entitled “Measuring costs and benefits of independence”, which was part of a joint project with Prof J.V. Rodriguez Mora and Mr Nicholas Myers.

This work has not been submitted for any other degree or professional qualification.

David Comerford

13th June 2013

A balance of questions: what can we ask of climate change economics?

Abstract

The standard approach to the economics of climate change, which has its best known implementation in Nordhaus's DICE and RICE models (well described in the Nordhaus (2008) book, *A Question of Balance*) is not well equipped to deal with the possibility of catastrophe, since we are unable to evaluate a risk averse representative agent's expected utility when there is any significant probability of zero consumption. Whilst other authors attempt to develop new tools with which to address these problems, the simple solution proposed in this paper is to ask a question that the currently available tools of climate change economics are capable of answering. Rather than having agents optimally choosing a path (that differs from the recommendations of climate scientists) within models which cannot capture the essential features of the problem, I argue that economic models should be used to determine the savings and investment paths which implement climate targets that have been suggested in the physical science literature.

1 Introduction

This article argues that there are two different questions that an economic analysis of climate change could address when providing advice to policymakers using dynamic general equilibrium models. These questions are: ‘*What is the social cost of carbon given a climate change externality?*’ and ‘*Given a cumulative emissions target, what carbon tax implements this target?*’ Both these questions are important, but in giving policy advice, we can provide a much less uncertain answer to the second question than we can to the first. The positive methodology employed in answering these two questions is the same but the normative question is different. The standard approach to climate change economics, as pioneered by Nordhaus (see Nordhaus (2008) for a comprehensive summary), attempts to answer the first of these questions by evaluating the price of carbon (equal to the social cost of carbon) that a social planner would have to set, in order to implement the policy programme that maximises their lifetime CRRA utility objective function, in a world where high temperatures damage production or utility, and production in the absence of abatement technology causes high temperatures. Answering the second question involves treating a given climate change target as a resource constraint, and evaluating the tax that incentivises investment in a zero carbon economy such that this investment is complete before the resource constraint is bust.

The paper is structured as follows: In section 2, I discuss the literature and describe the climate change policy advice that comes from both the climate-economy and climate science literatures, and highlight differences between them. In section 3, I classify economic modelling efforts according to the assumptions about catastrophe that implicitly lie behind them, and place the physical science policy advice within this framework. This classification exercise highlights the fundamental reasons for the differences in policy recommendations that are discussed in section 2, and motivates a focus on the second question: ‘*What is the carbon tax required to implement a scientifically determined target?*’. Section 4 discusses the fact that, given current knowledge, and the currently accepted welfare framework, the only rigorous answer to the first question, ‘*What is the social cost of carbon?*’, is undefined, which further motivates a focus on the second question. This section also describes some of the work that is being done on developing alternative welfare frameworks, which will eventually (but not yet) allow us to answer this first question. In section 5, I sketch a solution to the second question, and highlight the result that asking a different question can produce qualitatively different policy advice: I present a strong result on the implication of targeting a cumulative emissions limit on the level and timepath of carbon taxes that is not revealed by standard climate economy models. Section 6 concludes.

This paper therefore both makes clear the questions that are important in climate change economics, and provides an outline answer to the easier of the two. Despite being an easier question,

it is a less researched and possibly more policy relevant question. Given a super-abundance of fossil fuels, and a backstop technology that is only available with large scale investment, what carbon price path should the policymaker implement to meet a cumulative emissions target? Questions like this have been addressed before (e.g. Nordhaus (2008) talks of the application of the DICE 2007 model to binding temperature and CO_2 concentration limits) and it is the discussion of the prominence that such questions are given, and the implications of answering different questions, that is the contribution of this paper. The question that this paper claims is should be focused upon in climate change economics is consistent with the argument made by Sinn (2009) in which he says “the core question ... [is] how to induce the resource owners to leave more carbon underground, as that is the sole possible way to solve the climate problem.” However, I argue that the framework in which Sinn (2009) operates cannot address the problem he sets himself.

2 Climate change policy advice

The Intergovernmental Panel on Climate Change (IPCC, Solomon, Dahe, Manning, and et al. (2007)), stated and evidenced that there were “reasons for concern” that climate change greater than $2 - 3^\circ C$ may be dangerous. This “danger” is due to the possibility that there may be thresholds in the climate system that mean large changes, outwith the range of model predictions, are possible. Many climate scientists are coming to the belief that avoiding thresholds, or tipping points, in the climate system, is the crucial aspect of climate policy (see e.g. Alley, Marotzke, Nordhaus, Overpeck, Peteet, Jr., Pierrehumbert, Rhines, Stocker, Talley, and Wallace (2003), Overpeck and Cole (2006), & Lenton, Held, Kriegler, Hall, Lucht, Rahmstorf, and Schellnhuber (2008)). In 2005 the European Union adopted a $2^\circ C$ temperature rise limit (above pre-industrial global temperatures) as a policy goal. Given central estimates of climate sensitivity to increases in CO_2 concentrations of $\sim 3^\circ C$ for a doubling of atmospheric CO_2 , this implies a CO_2 concentrations limit of $\sim 450ppm$ (given that pre-industrial concentrations were $\sim 280ppm$). However, as evidence accumulates, some have argued that the $450ppm$ target is too lenient, e.g. Hansen, Sato, Kharecha, Beerling, Berner, Masson-Delmotte, Pagani, Raymo, Royer, and Zachos (2008) recommend a target of, and describe a scenario whereby, atmospheric CO_2 levels are down to no more that $350ppm$ by 2100.

A particularly easy to express and communicate target is introduced by Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen (2009) who note that climate models seem relatively insensitive to the timepath of emissions, and rather cumulative emission targets are much more useful. In keeping with the $2^\circ C$ limit, they suggest a cumulative emissions target of 1 trillion tonnes of carbon ($1TtC$). Given that historical emissions since the start of the industrial revolution are estimated at around 500 billion tonnes of carbon ($500GtC$), this means we are half

way through our allowance, and have $500GtC$ left to burn. This is much less than the available fossil fuel resource (see [Brandt and Farrell \(2007\)](#)).

This policy advice from the climate science community can be contrasted with that coming from the climate-economy literature. To summarise this, I focus on [Tol \(2009\)](#) which surveys prior literature. In surveying the economics of climate change field, Tol notes that an “area of agreement between these studies is that the welfare effect of a doubling of the atmospheric concentration of greenhouse gas emissions on the current economy is relatively small - a few percentage points of GDP. ... roughly equivalent to a year’s growth in the global economy - which suggests that over a century or so, the economic loss from climate change is not all that large.” This summary is borne out by the fact that the climate-economy studies do not deem it optimal to stick to a $2^{\circ}C$ temperature rise limit. This disagreement with the recommendations coming from the climate science community may seem strange given that many of these models include a simplified environmental model that has been calibrated to reproduce the temperature rises seen in climate models. For example, the DICE 2007 model described in [Nordhaus \(2008\)](#) has been calibrated to the MAGICC climate model¹. And the evidence for economic damages from a given level of warming, whilst highly uncertain, has been researched (see [Nordhaus \(2008\)](#)) and cannot be dismissed.

The disagreement in policy recommendations, ultimately arises from the possibility of catastrophe - from the possibility that the actual climate response to a large increase in CO_2 concentrations will be much greater than implied in the climate science models. Beyond $2 - 3^{\circ}C$ ‘here be dragons’ and explore these regions at your peril. The climate-economy models surveyed by [Tol \(2009\)](#) agree with the central projections from the climate science models by construction, and it is likely that climate damages associated with these projections are low in the context of overall global output - but this is because no tipping points are breached, and catastrophic dynamics are not set in motion. The calibration of the climate side of climate-economy models to the central projection from models used in climate science, whilst clearly constraining the policy recommendations from the climate-economy literature, is not capturing the state of scientific knowledge of climate change. Experience to date has tended to track the worst case simulations from climate science models, suggesting that these models have been constructed on a conservative basis ([Bryson, Oreskes, O’Reilly, and Oppenheimer \(2013\)](#)). Further, climate science models have difficulty in matching some features of the known paleoclimate record, and of matching the tipping-point, threshold behaviour, suggested by paleoclimatic data (see [Valdes \(2011\)](#)). The central IPCC climate projections contain no threshold effects at a $2^{\circ}C$ temperature rise and the accompanying warning of “reasons for concern” that climate change greater than $2 - 3^{\circ}C$ may be dangerous, is due to the

¹Model for the Assessment of Greenhouse-gas Induced Climate Change. See <http://www.cgd.ucar.edu/cas/wigley/magicc/>

sense that climate science practitioners have that the models' central projections cannot be trusted for large climate forcings - largely because of this failure to match the paleoclimate record.

In addition to the treatment of catastrophe causing a problem in deriving an optimal policy in climate-economy studies (discussed further in the next section), there is a further problem with surveying the social cost of carbon (or equivalently the optimal carbon tax) across the economics of climate change literature and producing mean or median estimates that may be regarded as the considered view of the economics profession. Many articles make a methodological contribution or add some new feature to the climate-economy models. Their social cost estimate should therefore not be viewed as a data point towards what the economics profession think the optimal carbon price is, but rather as determining the marginal value of this new feature on the optimal carbon tax. For example, [Lemoine and Traeger \(2012\)](#), and [Brock, Engstrom, and Xepapadeas \(2012\)](#) describe climate economy models with a non-catastrophic "tipping point" and so their contribution can be viewed as estimating the marginal value of this feature. [Golosov, Hassler, Krusell, and Tsyvinski \(2011\)](#) derive, under certain assumptions, a closed form optimal tax formulation that depends only on the parameters of the model rather than knowledge of the future evolution of the model's endogenous variables, and so contribute to the climate change economist's toolbox rather than providing a data point for a survey.

3 Classifying approaches to climate change economics

I classify climate-economy studies into 4 groups depending on whether there is or is not a tipping point, whether avoiding crossing this tipping point on a CO_2 stabilisation path is technologically feasible, and whether crossing the tipping point is catastrophic or not. Before describing this classification, I shall first outline exactly what I mean by a tipping point and discuss some evidence that means that we cannot rule out tipping points being catastrophic for global civilisation.

A tipping point is a feature of dynamical systems with multiple steady states. As some forcing (like CO_2 from human emissions, or insolation (a measure of solar energy received on the Earth's surface)) is applied to a system (like the Earth's climate), the equilibrium state of the system may change smoothly, or it may change discontinuously across a tipping point. Reversing the forcing change need not reverse the movement across the tipping point. For example [Zaliapin and Ghi \(2010\)](#) present a simple *energy balance* model of an water-world Earth-like system that is vulnerable to a catastrophic cooling to a *snowball Earth* state. The mechanism for this is that ice is much more reflective than water and so a world with intermediate levels of incoming solar radiation can be either cold or hot. If cold then the planet is ice-covered and heat reflecting (high albedo) which induces energy balance at a cold temperature, if warm then the planet is water-covered and heat absorbing (low albedo) which induces energy balance at a warm temperature i.e. two steady

states. Suppose the system is in the cold, ice-covered steady state, then in order to transition to the warm steady state, incoming solar radiation has to be raised to a very high level to the point where ice cannot exist, at which point the cold steady state cannot exist and temperatures rise catastrophically as we cross the *tipping point*. The situation is well described by the bifurcation diagram, figure 3.1, taken from [Zaliapin and Ghi \(2010\)](#) which shows the equilibrium insolation and temperature combinations in their model.

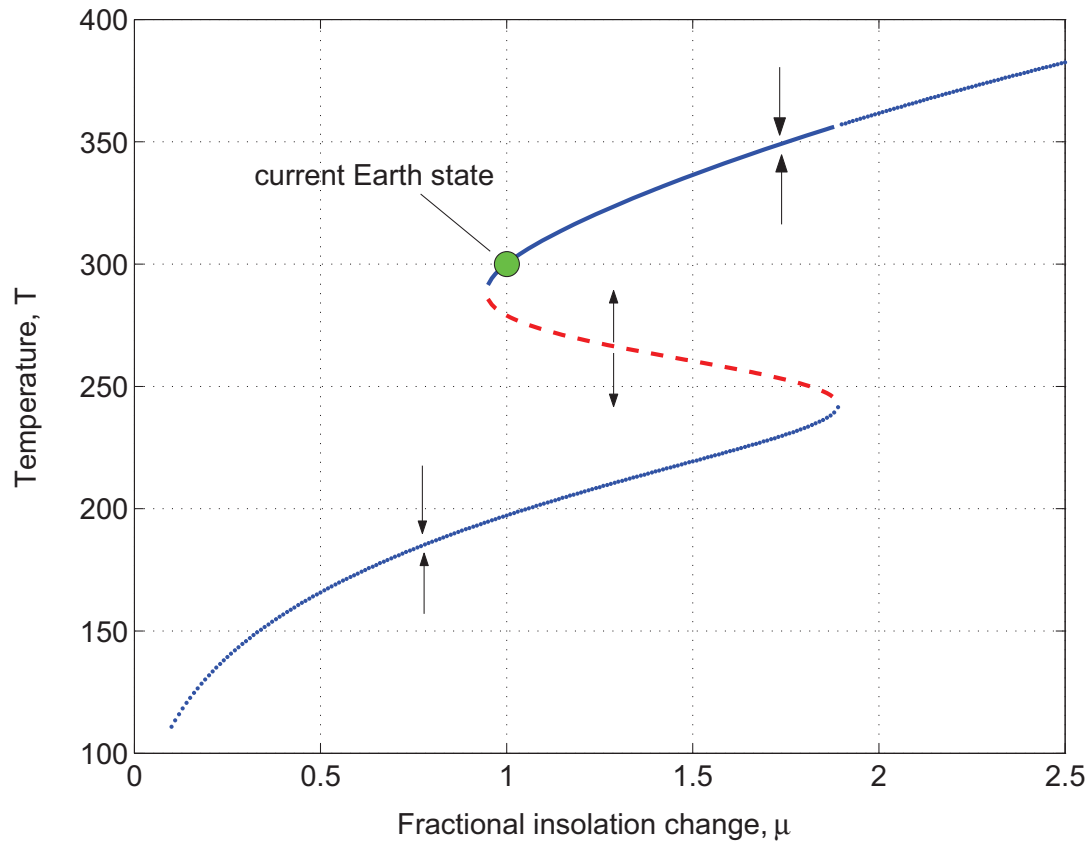


Figure 3.1: Energy Balance model of Snowball Earth from [Zaliapin and Ghi \(2010\)](#)

There are multiple sources of possible tipping points in the real Earth system including:

- The loss of polar ice at higher temperatures leading to a loss of albedo and a further rise in temperatures;
- Higher temperatures reducing ocean turnover, which reduces ocean productivity and carbon absorption and storage, causing a further rise in temperatures;
- Higher temperatures drying out peat bogs at high latitudes, melting methane hydrates in

ocean sediments, and burning tropical forests, all causing carbon emissions from the natural environment, and hence a further rise in temperatures.

Each of these could lead to tipping points as the Earth transitions from one steady state to another. And once we hit a tipping point, change can be very fast: we have evidence (Alley (1993)) that there were changes (at least at the regional, if not the global, level) of as much as 10°C in as little as a decade, as Earth was coming out of the last ice age. Given our industrial society and economy, such rapid change could be very destabilising and damaging.

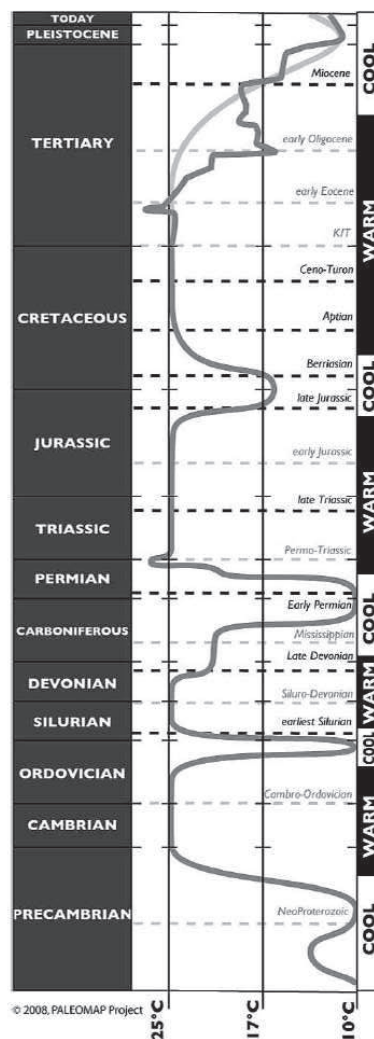


Figure 3.2: Approximate paleoclimate history, taken from <http://www.scotese.com/climate.htm>, created by Prof C Scotese, University of Texas at Arlington, Dept. Earth & Environmental Sciences

The geological record of global temperatures is suggestive of multiple steady states (see Figure 3.2). Further, there is evidence that the majority of the mass extinctions in the fossil record are

associated with greenhouse warming events (see [Ward \(2006\)](#) for a summary). It therefore seems clear that it is certainly possible that mechanisms exist such that sudden drastic warming is caused as a tipping point is passed, and that the impact of this warming is catastrophic - possibly enough to cause civilisational collapse or even human extinction. This may not be the central expectation, but many physical scientists are warning that catastrophe is possible: see [Hansen \(2009\)](#) for a comprehensive account of the fears of one prominent climate scientist who clearly believes that human caused greenhouse gas emissions could lead to an extinction level catastrophe.

The 4 groups in the climate-economy study classification are described below. The charts accompanying these classifications have been created with the simple climate-economy model with tipping point outlined in the Appendix.

1. No tipping point: damages, however severe, are a smooth function of stabilisation CO_2 concentrations, and are never catastrophic. This is the case considered by [Nordhaus \(2008\)](#), [Golosov, Hassler, Krusell, and Tsyvinski \(2011\)](#), and in the underlying papers of the [Tol \(2009\)](#) survey.
2. A non-catastrophic tipping point. The optimum may be before, at, or after the tipping point, and stabilisation before may or may not be feasible. This is the case that [Brock, Engstrom, and Xepapadeas \(2012\)](#) and [Lemoine and Traeger \(2012\)](#) model. Figures 3.3 and 3.4 illustrates the cases with feasible stabilisation before the tipping point, and with optima at and after the tipping point.

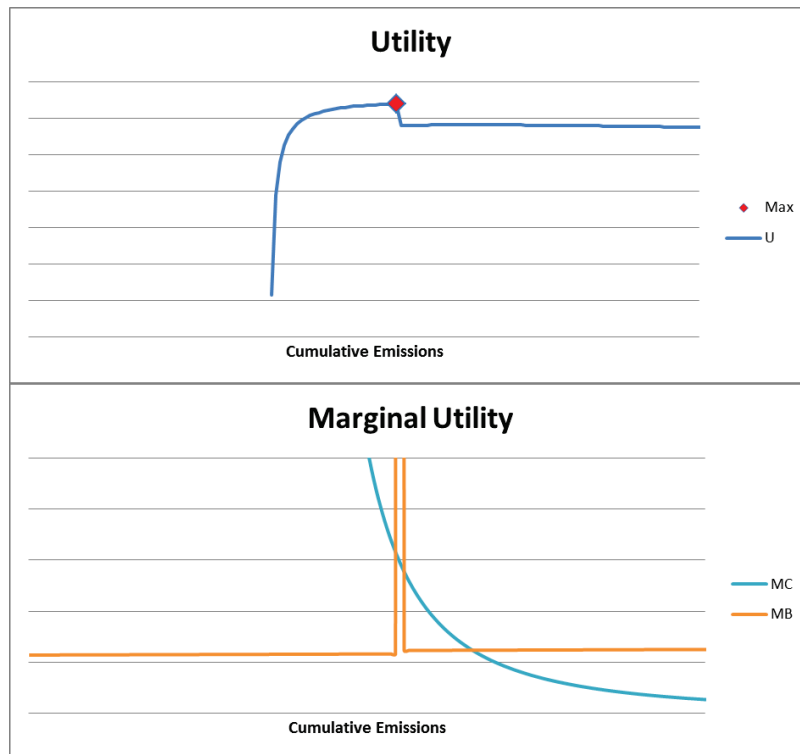


Figure 3.3: Utility (= Utility Benefit - Utility Cost) & Marginal Utility (with Marginal Benefit & Marginal Cost shown separately) with a non-catastrophic tipping point - stabilisation at the tipping point is optimal

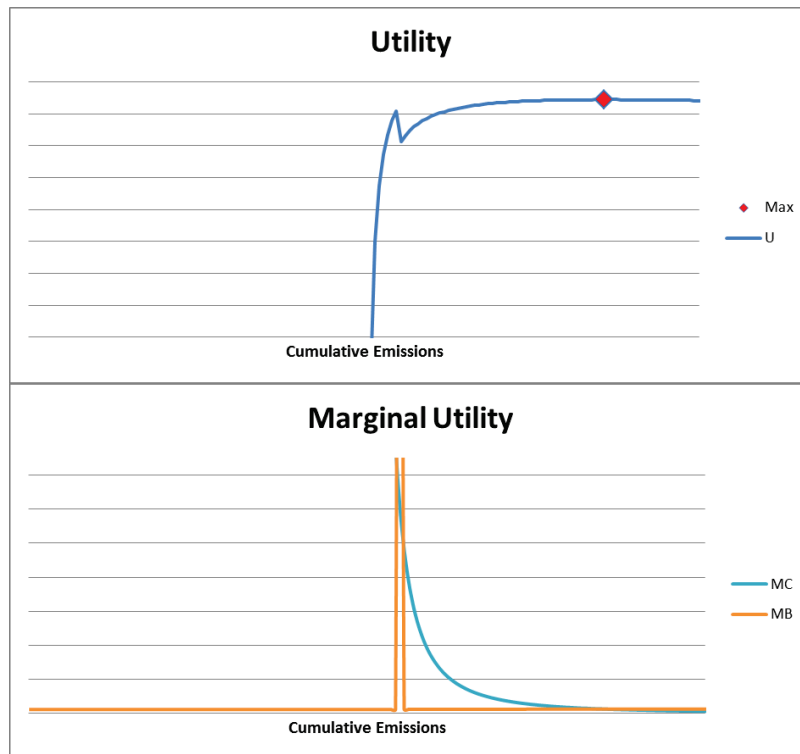


Figure 3.4: Utility & Marginal Utility with a non-catastrophic tipping point - stabilisation after the tipping point is optimal

3. There is a case where we cannot afford to stabilise before the tipping point (since the costs of doing this are greater than output in the policy period), but crossing the tipping point is catastrophic and corresponds to some civilisational collapse or extinction event as discussed above. This maximisation problem has no solution. Considering this to be one possible outcome in an *ex-ante* unknown problem, then any significant probability attached to this event will cause the optimisation under uncertainty to have no solution. This essentially is the case described by the *Dismal Theorem* in Weitzman (2009) (discussed further in Section 4): whatever the level of unaffordable stabilisation costs, there is some small but non-zero probability (fat-tailed density) that a catastrophe occurs before this stabilisation level.
4. A catastrophic tipping point, but stabilisation before this tipping point is possible. There is infinite marginal benefit of climate policy implementation at the tipping point. Figures 3.5 and 3.6 illustrates this. This is the scenario that the climate science community is warning about: we cannot rule out tipping points being catastrophic, but it is likely that we can do something about the problem if we act quickly and decisively to implement climate policy. Stabilisation at or before the tipping point is optimal. Given uncertainty about exactly where the tipping point is, stabilisation exactly at the tipping point will be impossible. Assuming

scientific advice to be erring on the cautionary side, following scientific advice is stabilisation before the tipping point. This is a realistic policy scenario.

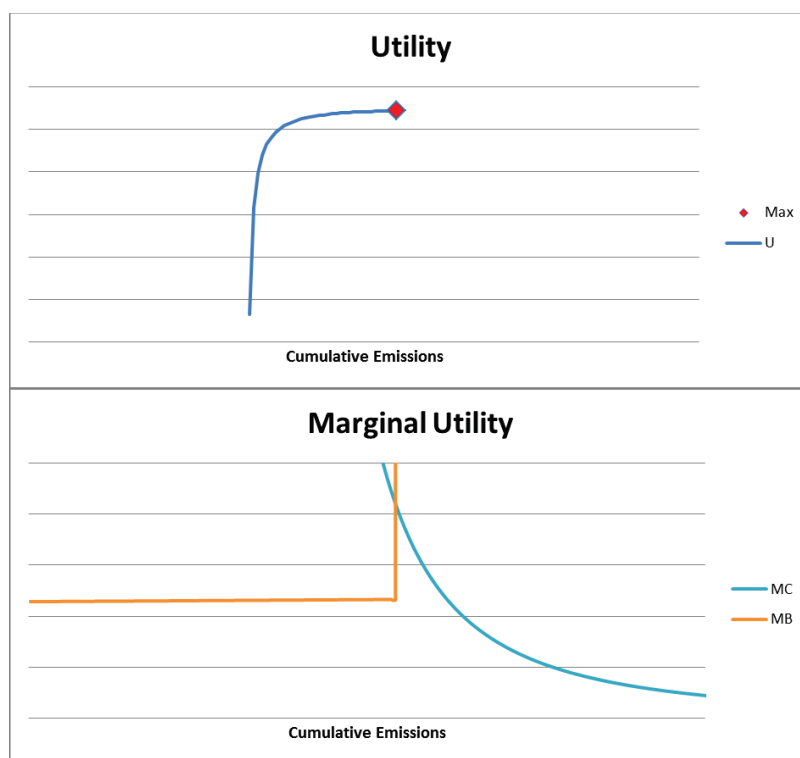


Figure 3.5: Utility & Marginal Utility with a catastrophic tipping point - stabilisation at the tipping point is optimal

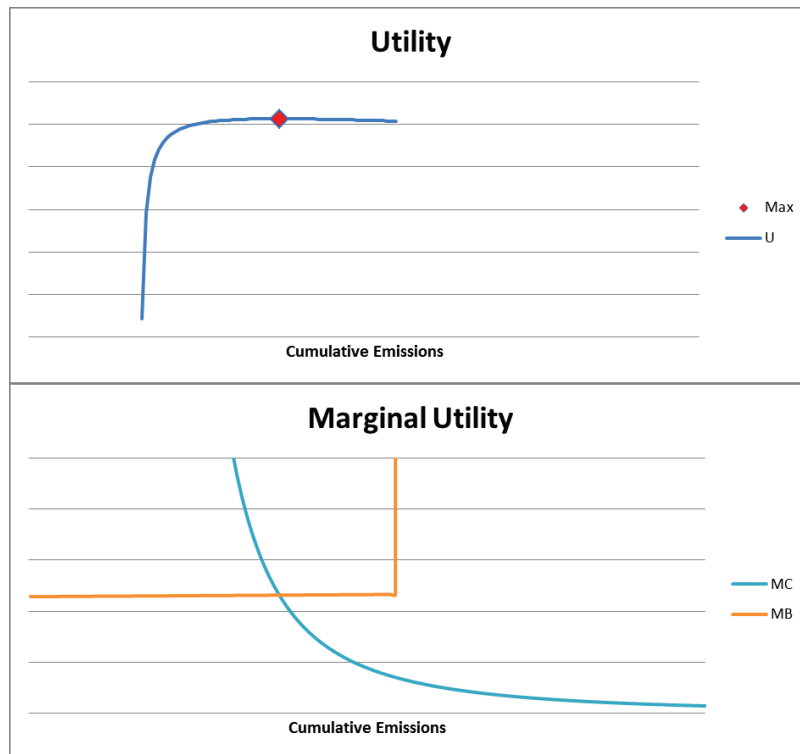


Figure 3.6: Utility & Marginal Utility with a catastrophic tipping point - stabilisation before the tipping point is optimal

Given this classification, we can see clearly where the divergence in policy advice between those arguing within the paradigm of case 4 (climate science recommendations), and those arguing within the paradigm of case 1 (the majority of climate change economics), comes from. However, this classification also reveals that the work of Brock, Engstrom, and Xepapadeas (2012) and Lemoine and Traeger (2012), (case 2), does not help in bridging this divide: it may be optimal to stabilise at a point after a non-catastrophic tipping point, and Brock, Engstrom, and Xepapadeas (2012) and Lemoine and Traeger (2012) merely discuss particulars of such models. The existence of non-catastrophic tipping points does affect optimising behaviour in ways related to the system dynamics and the degree of risk aversion of the agents, however these models do not help us at all in determining where the tipping points are, or how damaging they will be. Given that we cannot rule out catastrophic damages (and indeed suspect that they may exist), then unless we change the welfare framework (discussed in Section 4), we are, at least in a probabilistic sense, in case 3 (cue tearing of hair and gnashing of teeth, we're all doomed!) or case 4 (sensible but urgent policy advice). Stabilisation before any tipping point is likely (though not certainly) feasible and optimal. Climate models, being detailed models of the climate, are more appropriate than any climate-economy models (which necessarily have more approximation in order to facilitate optimisation) for determining where the tipping points are. These climate models have not answered this question

yet, so the best we, as economists, can do is to trust what practitioners in this field say and recommend. For the economics of climate change, as Pindyck (2012) says “it seems to me that a very detailed and complex modeling exercise is unlikely to be helpful”, we should let the climate science community do the heavy modelling work, take their recommendations, and concentrate on modelling the economic impacts and determining the economic instruments that implement policy.

4 Dealing with the *First Question*

Nordhaus’s standard approach to climate change economics, which (as discussed) is an effort to answer the ‘*What is the social cost of carbon?*’ question, was extensively criticised by Weitzman in a series of papers (see principally Weitzman (2009)) in which he shows that, allowing for uncertainty, this carbon price is infinite. Weitzman’s *Dismal Theorem* applies to problems which use an objective function with infinite marginal utility of consumption at the zero consumption level, combined with an effective probability mass² attached to catastrophe, which equates to this zero consumption level.

It is easy to object to this conclusion, which implies that society would optimally allocate an arbitrarily high share of current output to preventing tiny but non-zero risks of catastrophe or extinction. And while we may believe that society should devote more resources to climate change mitigation (or to developing asteroid or super-volcano protection systems etc) than it currently does, we are unlikely to believe that the resources so allocated should be approximately 100% of current output. However, on what grounds do we object to Weitzman’s theorem? Which of the underlying assumptions do we disagree with?

According to Millner (2011), there are three bottom-up grounds to object to Weitzman’s methods and, therefore, conclusions: is it reasonable to attach a probability mass point to the possibility of catastrophe; is the assumption of infinite marginal utility of consumption at a zero consumption level reasonable; and, is this analysis under uncertainty relevant to a Cost Benefit Analysis (CBA) of climate policy? Millner concludes that only the infinite marginal utility of consumption critique has validity. Therefore, in order to evaluate the social cost of carbon we need to study how to (finitely) value catastrophe risks i.e. risks to civilisation or risks of human extinction.

Weitzman (2009) discusses doing this by truncating the valuation attached to bad events, using a Value of Statistical Life (VSL) method. However this approach is shown to generate results in which the truncation becomes the dominant factor in the CBA calculation (so the size of the median impact does not really effect the calculation, the impact comes almost entirely from the choice of

²Weitzman’s result is stronger than this in that it actually just relies on fat tails, either in the distribution of environmental outcomes or in the distribution of economic damages associated with a particular level of environmental outcome. Describing this result as applying due to a probability mass on catastrophic outcomes is intuitive though.

truncation methodology). Ikefuji, Laeven, Magnus, and Muris (2011) describe a stochastic climate economy model with non CRRA utility specifications chosen to produce robust policy prescriptions. This is an attempt to deduce a welfare framework on decision making under uncertainty about catastrophe from the axiom that ex-ante policy is both optimal and *reasonable* (i.e. not 100% of output). Millner (2011) is also a discussion the development of a welfare framework in which it is sensible to ask how we should value civilisation preserving policies, and which does not run into the Dismal Theorem’s paradoxical infinities. This question is an interesting and important research question, but it is also a hard problem (and perhaps more philosophical than economic), and in the meantime, what advice do we give to policymakers on optimal climate change policy?

The answer proposed in this paper is to change the question: if we assume that following the advice of the climate science community eliminates the risk of catastrophe, then we could use a CRRA utility specification that exhibits infinite marginal utility at zero consumption, since we would only be using this welfare framework in its *natural setting* i.e. for consumption-savings decisions well away from the zero consumption level. We cannot strictly make this assumption, therefore we are not asking ‘what is the optimal policy?’, rather we are imposing scientific advice as a resource constraint and asking what price implements this constraint. The normative question is different - but still policy relevant and can be reconciled with the views of the climate science community.

5 Sketching a solution to the *Second Question*

In this section, I set out to answer the second question that practitioners within the economics of climate change should be addressing: ‘*Given a cumulative emissions target, what carbon tax implements this target?*’. I base the model I use to answer this question, to a limited extent, on Golosov, Hassler, Krusell, and Tsyvinski (2011), not because I intend to argue with this paper in particular, but because it is a recent paper with a clear calibration that I can use. However, the main purpose of showing this exercise is to highlight the features that need to be included within such a model (and which are often not included in climate-economy models that attempt to answer the first question) and including these features has a strong impact upon the resulting policy advice. The model in Golosov, Hassler, Krusell, and Tsyvinski (2011) does not have all these required features and the policy recommendation from this paper repeat the claims of Sinn (2009) that ad-valorem taxes on fossil fuels should fall over time in order to prevent resource owners bringing forward their resource extraction activities (the so-called *Green Paradox*). I present different implications for the level and timepath of taxes - this highlights the fact that asking different questions can produce different answers!

Sinn (2009) states that “there are only two ways to curb the accumulation of carbon dioxide in the atmosphere and, with it, slow down global warming. We either temporarily refrain from extracting carbon from the ground, or we stuff it back into the ground after having extracted its energy. All the technical endeavours to develop alternative technologies and all economic incentive systems to curb the greenhouse effect must subordinate themselves to this fundamental fact” and “carbon extraction rates must be slowed down. The resource owners must be prompted to temporarily leave more carbon underground.” Clearly, Sinn (2009) does not consider a cumulative emissions limit³, but rather considers only managing the timepath of carbon emissions. The framework that he considers then is already inconsistent with the findings of Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen (2009). He further argues that it is only a political economy issue that determines why policymakers might choose to have a low level of climate policy now which is progressively made more stringent.

The theoretical solution to the Green Paradox presented by Sinn (2009) is to have stringent climate policy which becomes more lax over time: a falling ad-valorem tax. This would incentivise resource owners to postpone extraction. Golosov, Hassler, Krusell, and Tsyvinski (2011) also report that carbon taxes should fall over time and that constant ad-valorem taxes have no effect on usage. However, this conclusion is due to Inada conditions on the use of energy together with no available alternative energy technologies (which are understandable modelling simplification devices if the study is to address the question of ‘*What is the optimal carbon tax given a climate change externality?*’). Given the Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen (2009) conclusion that cumulative emissions are what matters and that some fossil fuels should be left in the ground, making these assumptions mean that the conclusions of these papers are not capable of being consistent with the climate science advice. By asking a different question, ‘*Given a cumulative emissions target, what carbon tax implements this target?*’, the model presented here reaches a different conclusion on the time path of carbon taxes: ad-valorem taxes must always be set at 100% in order for the owners of zero marginal cost resources to be indifferent between extraction and leaving resources in the ground. A 100% ad-valorem tax is a per unit energy tax equal to the marginal product of energy. If we are to use a lot of energy now, in order to create an alternative energy infrastructure, then the marginal product of energy should be low now and rise to some constrained energy future steady state i.e. we see a rising per unit tax on carbon. This shows that, if we ask different questions, we get different answers.

Golosov, Hassler, Krusell, and Tsyvinski (2011) is a standard, smooth damages, climate economy general equilibrium model that has a number of interesting features, but whose main contribution is the derivation, under certain assumptions, of a closed form optimal tax formulation that

³He does state that “the argument for permanently sealing off part of the resources still in situ to the detriment of generations far in the future finds neither economic nor ethical justification.” But he does not justify this assertion.

depends only on the parameters of the model rather than knowledge of the future evolution of the model's endogenous variables. The *optimum* level of carbon emissions in their results is much greater than the 500GtC future cumulative emissions limit that [Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen \(2009\)](#) recommend. Their optimal policy is determined by balancing the marginal costs of climate change against marginal benefits of energy use (non carbon energy sources are not available in their model until after the model's time horizon of more than a century). Their model does not consider the development of, and investment in, the alternative energy technologies needed for the post fossil fuel world. The focus therefore of the [Golosov, Hassler, Krusell, and Tsyvinski \(2011\)](#) model (in common with most such models) is on the environmental side which, as previously argued, climate economy models are poorly equipped to deal with, and not at all on the investment side which economic models have comparative advantage in addressing.

I develop a simple two energy sector economy model with an available stock of carbon energy resources, and also the technological possibility of building non carbon energy infrastructure. The model is calibrated so that the social planner, faced with a resource constraint of using only those resources used in the [Golosov, Hassler, Krusell, and Tsyvinski \(2011\)](#) optimum, sets a carbon tax equal to that derived in the initial period of [Golosov, Hassler, Krusell, and Tsyvinski \(2011\)](#) i.e. \$56.9/tC. Using this calibration, I can then estimate the carbon tax needed to implement a resource restriction consistent with [Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen \(2009\)](#). Climate damages do not enter the model: if we were to include climate damages that were very low before the tipping point but infinite after it, instead of implementing this via a resource constraint, then the results would not be much altered. The details of the model and the solution algorithm are outlined in Appendix 2, but broadly the social planner has logarithmic preferences over consumption, energy is essential for production and energy itself can be produced from some stock S_0 of fossil fuel or from some *green-tech energy capital* stock. The social planner's problem is to maximum lifetime utility by choosing consumption, investment in capital, and investment in green-tech energy capital, subject to the available fossil fuel energy resources.

We can determine the social planner's optimum solution to the saving and investment problem given various different values for S_0 . There is some *true* value for S_0 (i.e. the S_0 that pertains in a laissez-faire world) but we want the decentralised solution to mimic social planner's solution from the constrained S_0 run. We suppose there are a unit mass of identical, infinitesimal, price taking, profit maximising fossil fuel resource owners whose extraction costs are always zero but whose stock is limited to S_0 . Their profit maximisation decision is the quantity of fossil fuel, F_t to

supply in each period:

$$\begin{aligned} V_t(S_t) &= p_t F_t + r_{t+1} V_{t+1}(S_{t+1}) \\ \text{s.t. } S_{t+1} &= S_t - F_t \end{aligned}$$

This problem has solution:

$$p_t = r_{t+1} V'_{t+1}(S_{t+1}) = V'_t(S_t)$$

Given that we want to impose a cumulative emissions limit $S^* \ll S_0$, there must be some $S_T = S_0 - S^*$ for which the marginal value of the stock (after the imposition of policy) is zero. But this implies that the marginal value in all periods prior to T is also zero, and hence that the net price receive for any resources extracted, $p_t, \forall t$, is also zero. The optimal carbon tax then is equal to the marginal product of energy in the social planner's solution from the constrained S_0 run, and fossil fuel suppliers are indifferent about the level of fuel that they supply. In equilibrium the fossil fuel supply in this world with super-abundance of fossil fuel resources is the same as that achieved in the social planners solution with the constrained S_0 .

The carbon tax must be equal to the marginal product of energy so that the final goods sector's net payment to the fossil fuel industry is always zero. The value of fossil fuel resources to their owners is therefore always zero too (though the marginal value to society of increasing the amount of fossil fuels that we are willing to burn is most definitely not zero). We can therefore implement a restriction on the total allowable burnable fossil resource even in the presence of a super-abundance. At time T the energy sector is entirely decarbonised. We proceed as follows (again full details in Appendix 2):

- Assume the calibration of [Goloso, Hassler, Krusell, and Tsyvinski \(2011\)](#) and calibrate production function so that current global capital stock and energy usage produces current global GDP.
- Use the [Goloso, Hassler, Krusell, and Tsyvinski \(2011\)](#) carbon budget (substantially greater than $500GtC$) under their *optimum policy* as a resource constraint, and use their initial carbon tax of $\$56.9/tC$ as a further calibration target to fully calibrate the model.
- Use the [Goloso, Hassler, Krusell, and Tsyvinski \(2011\)](#) carbon budget under their *no policy* as a resource constraint and label as “Laissez Faire” in Figures 5.1, 5.2 & 5.3.
- The policy we cost here, from [Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen \(2009\)](#), is restricting S_0 to $500GtC$. This model run is labelled as “ $500GtC$ ” in Figures 5.1, 5.2 & 5.3.
- The required carbon tax in the $500GtC$ policy scenario is found to be $\$70.5/tC$ at the outset.

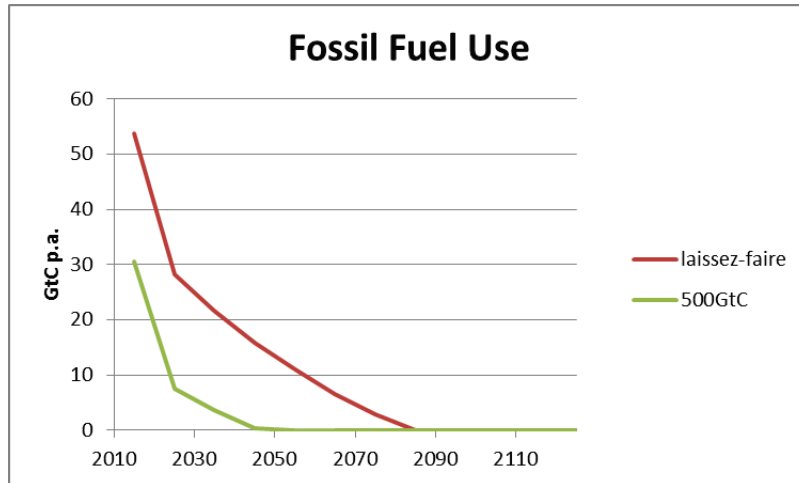


Figure 5.1: Fossil fuel use, in $1GtC$, vs time

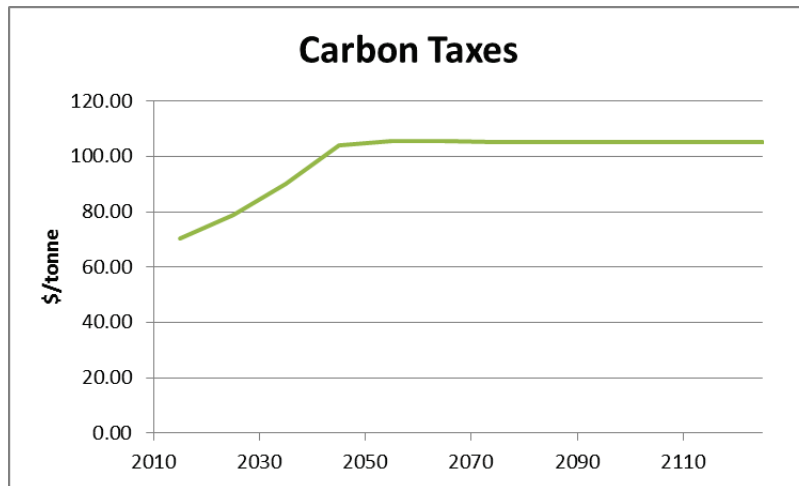


Figure 5.2: Carbon taxes, in $\$/tC$, vs time

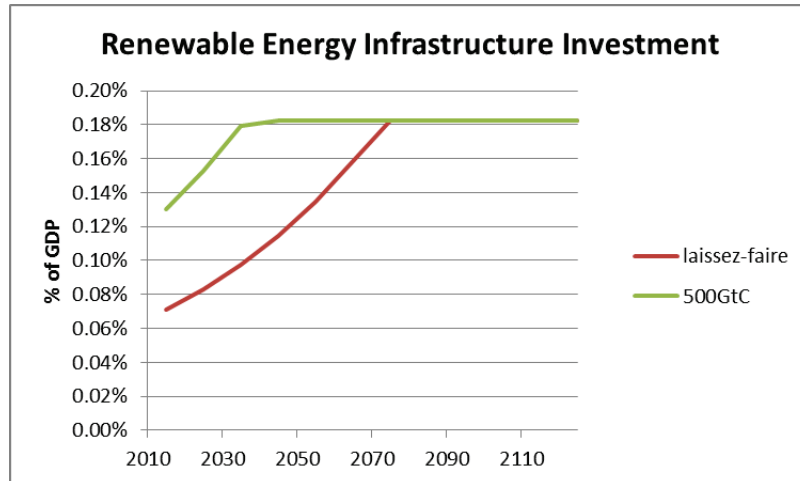


Figure 5.3: Investment in alternative energy capital, as % of GDP , vs time

As well as reporting a higher tax needed to implement an emissions target that we already know to be tighter than the optimum reported in [Goloso, Hassler, Krusell, and Tsyvinski \(2011\)](#) (not a surprising result), it is also interesting to report the time path of taxes and resource usage. As discussed, in order to induce resource owners to leave resources in the ground when marginal extraction costs are zero, the value of extraction must always be zero. This means that ad-valorem taxes are 100% and dollar taxes are equal to the marginal product of energy. At outset, the planner would like to use the endowment of allowed fossil fuels to produce and so consume whilst also investing in alternative energy infrastructure. In a world where we are relatively energy constrained in the no-fossil fuels future compared with today, the ultimate carbon tax will be fairly high (because the marginal product of energy will be high). Initially there is relative energy abundance and the marginal product of energy (which is the price that the final goods sector pays for its energy), is lower than it will be in long run steady state (which relies entirely on the backstop technology). Therefore, the dollar value of carbon taxes will be rising. Contrary to [Sinn \(2009\)](#), it is not for political economy reasons⁴ that taxes are lower initially and rising, rather it is that marginal product of energy is low and rising, and this is because we are willing to use fossil fuels (lowering the marginal product of energy) to produce and invest in the early stages. We want to do this because at this time we have no alternatives, and these alternatives must be built with today's output.

The claim that “A constant value-added tax does not affect the intertemporal decisions of the firm, and hence has no effect on allocations, no matter how high this tax is” ([Goloso, Hassler,](#)

⁴[Sinn \(2009\)](#) “The largest reduction efforts are to be made in the far future, while the current generations are largely spared. Politicians cannot do otherwise, alas, as they do not want to inflict the pain of immediate reductions upon their voters. The year 2050 is so far in the future that the boldest policy proposals can be made now without scaring voters off. After all, the onus will fall later on other citizens and other politicians who will have to tighten their belts.”

Krusell, and Tsyvinski (2011)) is true only in a model in which there is no alternative energy supply (which does not pay the tax) and in which the purpose of policy is to manage the time profile of emissions rather than the cumulative total of emissions. The value added tax here is constant (100%) and this is a general result for all cumulative emissions targets with zero extraction costs. The fact that the dollar value of the carbon price is rising in the result generated here is not a general result and will depend on the comparison between the initial marginal product of energy in the social planner's solution, and the ultimate steady state marginal product of (carbon free) energy. The level of long run zero carbon energy supply is a key determinant of this steady state carbon tax level.

To fully answer the question of what is the optimal carbon price to implement a cumulative emissions target, we should include a more complete model of the energy sector. Such a model would include features such as fossil fuel extraction costs, inelastic demand for energy, and technological progress. A full study would build upon the work of e.g. Hassler, Krusell, and Olovsson (2011) who provide evidence that energy saving technical progress does respond to the energy price, and that the elasticity of substitution between energy and other factors of production is substantially less than 1; and P. and van Zon A. (2012) which is an example of a transition model which incorporates endogenous growth. I believe the study of the transformation and decarbonisation of the energy sector (and the interaction that this transition with the rest of the economy) is the central question of climate change economics, rather than the trade off between climate damages and the benefit flow from emitting carbon.

6 Conclusion

This article has argued that the reason for the dichotomy in policy advice between the climate science and economics literatures is fundamentally due to the treatment of catastrophic outcomes. To proceed, we can:

1. ignore the possibility of catastrophe, as much of the economics literature has done, and make policy recommendations that are far too light;
2. allow for catastrophe within our current welfare framework and recommend that we devote 100% of output to climate change mitigation (and another 100% of output to prevent other unlikely catastrophes!);
3. develop a new welfare framework in which we can sensibly evaluate policies that payoff only in preventing civilisational collapse or human extinction, and in the meantime, try not to muddy the waters with our half formed views of what appropriate policy is for climate change mitigation;

4. take the cumulative emissions limits, given to us by the climate science community, as resource constraints, and evaluate optimal policy conditional on staying within these constraints. The only way for the answer to this question to also be the answer to the “what is the optimal policy” question, is for scientists to be able to guarantee that following their advice would prevent catastrophe. Of course no such guarantees can be offered.

Whilst both options 3 and 4 above are sensible, this article is an appeal for more efforts to be put into 4. The minimal features that a model which can address 4 must exhibit are an alternative energy infrastructure technology and an ability for agents to leave resources in the ground. Such models could be used to provide highly relevant input that could inform the debate as to optimal climate change mitigation strategies.

References

- ALLEN, M., D. FRAME, C. HUNTINGFORD, C. JONES, J. LOWE, M. MEINSHAUSEN, AND N. MEINSHAUSEN (2009): “Warming caused by cumulative carbon emissions towards the trillionth tonne,” *Nature*.
- ALLEY, R. (1993): “Abrupt accumulation increase at the Younger Dryas termination in the GISP2 ice core,” *Nature*.
- ALLEY, R., J. MAROTZKE, W. NORDHAUS, J. OVERPECK, D. PETEET, R. P. JR., R. PIERREHUMBERT, P. RHINES, T. STOCKER, L. TALLEY, AND J. WALLACE (2003): “Abrupt Climate Change,” *Science*.
- BRANDT, A., AND A. FARRELL (2007): “Scraping the bottom of the barrel: greenhouse gas emission consequences of a transition to low-quality and synthetic petroleum resources,” *Climatic Change*.
- BROCK, W., G. ENGSTROM, AND A. XEPAPADEAS (2012): “Energy balance climate models, damage reservoirs and the time profile of climate change policy,” *Fondazione Eni Enrico Mattei Working Paper Series*.
- BRYSSA, K., N. ORESKESB, J. O’REILLY, AND M. OPPENHEIMER (2013): “Climate change prediction: Erring on the side of least drama?,” *Global Environmental Change*.
- GOLOSOV, M., J. HASSLER, P. KRUSELL, AND A. TSYVINSKI (2011): “Optimal taxes on fossil fuel in general equilibrium,” *NBER Working Paper Series*.
- HANSEN, J. (2009): *Storms of My Grandchildren*. Bloomsbury.
- HANSEN, J., M. SATO, P. KHARECHA, D. BEERLING, R. BERNER, V. MASSON-DELMOTTE, M. PAGANI, M. RAYMO, D. ROYER, AND J. ZACHOS (2008): “Target Atmospheric CO_2 : Where Should Humanity Aim?,” *Open Atmos. Sci. J.*
- HASSLER, J., P. KRUSELL, AND C. OLOVSSON (2011): “Directed energy-saving technical change,” *Preliminary*.
- IKEFUJI, M., R. LAEVEN, J. MAGNUS, AND C. MURIS (2011): “Weitzman meets Nordhaus: Expected utility and catastrophic risk in a stochastic, economy-climate model,” *The Institute of Social and Economic Research, Osaka University, Discussion Paper No. 825*.
- LEMOINE, D., AND C. TRAEGER (2012): “Tipping points and ambiguity in the economics of climate change,” *NBER Working Paper Series*.

- LENTON, T., H. HELD, E. KRIEGLER, J. HALL, W. LUCHT, S. RAHMSTORF, AND H. SCHELLNHUBER (2008): "Tipping elements in the Earth's climate system," *Proceedings of the National Academy of Sciences*.
- MILLNER, A. (2011): "On welfare frameworks and catastrophic climate risks," *NBER Working Paper*.
- NEHRU, V., AND A. DHARESHWAR (1993): "A New Database on Physical Capital Stock: Sources, Methodology and Results," *Revista de Analisis Economica*.
- NORDHAUS, W. D. (2008): *A Question of Balance*. Yale University Press.
- OVERPECK, J., AND J. COLE (2006): "Abrupt change in earth's climate system," *Annual Review of Environment and Resources*.
- P., D., AND VAN ZON A. (2012): "A multi-phase model of optimal transition to global climate stabilization," *Preliminary*.
- PINDYCK, R. (2012): "The climate policy dilemma," *NBER Working Paper Series*.
- SINN, H. (2009): "The green paradox," *CESifo Forum 3*.
- SOLOMON, S., Q. DAHE, M. MANNING, AND ET AL. (2007): *Intergovernmental Panel on Climate Change (IPCC), Climate Change 2007*. Cambridge Univ Press.
- TOL, R. (2009): "The Economic Effects of Climate Change," *Journal of Economic Perspectives*.
- VALDES, P. (2011): "Built for stability," *Nature Geoscience*.
- WARD, P. (2006): "Impact from the Deep," *Scientific American*.
- WEITZMAN, M. L. (2009): "On Modeling and Interpreting the Economics of Catastrophic Climate Change," *The Review of Economics and Statistics*.
- ZALIAPIN, I., AND M. GHI (2010): "Another look at climate sensitivity," *Nonlinear Processes in Geophysics*.

A Appendix 1

The very simple climate-economy model presented in this appendix is used to generate the utility and marginal utility charts in Section 3. The tipping point is generated with a methane hydrate reservoir of size \bar{M} with an emission process given by:

$$\begin{aligned}\dot{M}_t &= -\max[0, T_t - T_M^*]M_t \\ M_t &\in [0, \bar{M}]\end{aligned}$$

Where T_M^* is the destabilisation temperature for the methane hydrate stocks. There are two possible stable states for this system:

- If human emissions cause a temperature rise of less than T_M^* then methane hydrate stocks are not destabilised and ultimate CO_2 levels are just given by pre-industrial levels and those human emissions that remain in the atmosphere.
- If human emissions cause a temperature rise of more than T_M^* then methane hydrate stocks are destabilised and ultimate CO_2 levels are given by pre-industrial levels, remaining atmospheric human emissions and \bar{M} .

Figure A.1 illustrates this model given a particular parameter set. It shows equilibrium CO_2 levels against pre-industrial plus remaining atmospheric human emissions, generated using a logarithmic relationship between temperature change and CO_2 increases with a climate sensitivity of $3^\circ C$ for a doubling of CO_2 ⁵, $T_M^* = 3^\circ C$ (above pre-industrial temperatures), and $\bar{M} = 500ppm$ (which is assumed either to all remain in the atmosphere or to be the remaining atmospheric amount from the methane hydrate reservoir).

⁵This fixes the climate sensitivity parameter, $\kappa = (3^\circ C)/\ln(2)$

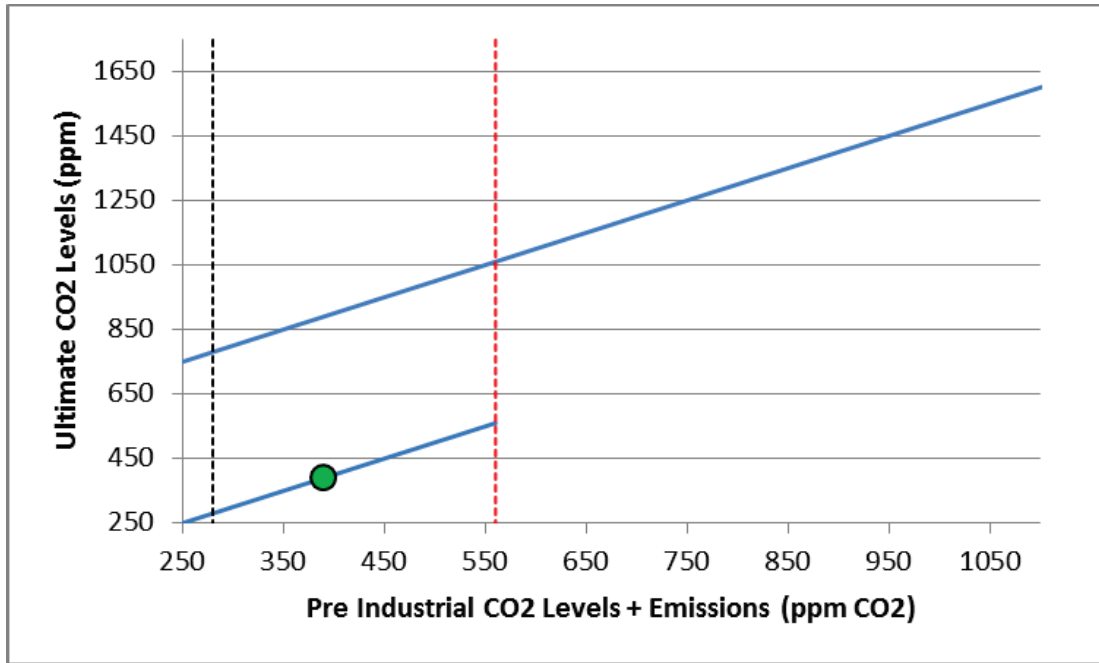


Figure A.1: Simple Tipping Point Model with Methane Hydrates

Using this model (though not necessarily the above parameterisation), we can impose a modified version of the climate damages function from Golosov, Hassler, Krusell, and Tsyvinski (2011)⁶. The Golosov, Hassler, Krusell, and Tsyvinski (2011) damage function is $D(S) = 1 - \exp(-\gamma(S - \bar{S}))$, with $\bar{S} = 280ppm$ being pre-industrial atmospheric CO_2 concentrations, and D expressed as the percentage of output lost due to high temperatures. This specification is modified to allow for the possibility that the tipping point is catastrophic i.e. in some circumstances we set $D(S > S_{tp}) = 100\%$. Economic output is assumed to be constant and entirely consumed, except in the first period when some is spent on climate policy, such that costs are hyperbolic in the stabilisation level. Utility is assumed to be CRRA, and climate policy is undertaken in the first period to achieve some stabilisation level of atmospheric CO_2 concentrations. This first period is assumed to be long enough to fully implement emissions elimination at some stabilised level of atmospheric CO_2 concentrations, and for the climate system to reach equilibrium. Utility as a function of the target CO_2 stabilisation level (S , expressed in $ppm CO_2$), and the other equations

⁶Which was calibrated to reproduce the damages from the DICE 2007 model described in Nordhaus (2008)

of this simple model are:

$$U(S) = \frac{(Y - C(S))^{1-\theta} - 1}{1-\theta} + \sum_{t=1}^{\infty} \frac{(Y(1 - D(S)))^{1-\theta} - 1}{1-\theta} \left(\frac{1}{1+r}\right)^t$$

where, $C(S) = \frac{A}{S - 394}$, $S > 394$

$$D(S) = \begin{cases} D_{NC}(S_{ult}(S)) & \text{if the tipping point is not catastrophic,} \\ D_{Cat}(S_{ult}(S)) & \text{if the tipping point is catastrophic.} \end{cases}$$

$$D_{NC}(S_{ult}(S)) = 1 - \exp(-\gamma(S_{ult}(S) - \bar{S}))$$

$$D_{Cat}(S_{ult}(S)) = \begin{cases} D_{NC}(S_{ult}(S)) & \text{if } \kappa \ln\left(\frac{S}{280}\right) < T_M^*, \\ 100\% & \text{if } \kappa \ln\left(\frac{S}{280}\right) > T_M^*. \end{cases}$$

and, $S_{ult}(S) = \begin{cases} S & \text{if } \kappa \ln\left(\frac{S}{280}\right) < T_M^*, \\ S + \bar{M} & \text{if } \kappa \ln\left(\frac{S}{280}\right) > T_M^*. \end{cases}$

We can represent the utility maximisation problem as a comparison between the marginal costs of implementing climate policy and the marginal benefits of implementing climate policy. Targetting CO_2 stabilisation at the tipping point S_{tp} (defined by $T(S_{tp}) = T_M^*$) is always (at least a one sided local) utility maximum and at this point there is a singularity in the marginal benefit of implementing climate policy.

B Appendix 2

Details and solution methods of the model used in Section 5. Variable definitions:

- $Y_t \equiv$ Output
- $C_t \equiv$ Consumption
- $L \equiv$ Population (assumed constant)
- $E_t \equiv$ Energy
- $R_t \equiv$ Renewable energy infrastructure
- $F_t \equiv$ Fossil fuels used
- $K_t \equiv$ Other physical capital
- $S_t \equiv$ Stocks of fossil fuels
- $I_t \equiv$ Investment in renewable energy infrastructure
- $0 < \alpha < 1 \equiv$ Capital share of total income
- $0 < \nu < 1 \equiv$ Energy share of total income
- $0 < \gamma < 1 \equiv$ Returns to scale in renewable energy production
- $0 < \beta < 1 \equiv$ Discount Factor
- $0 < \delta < 1 \equiv$ Depreciation rate (of both R & K)
- $A' \equiv$ Total factor productivity (assumed constant)
- $A = A'L^{1-\alpha-\nu}$
- $B \equiv$ Renewable energy infrastructure productivity (assumed constant)

Technologies and resource constraints:

$$\begin{aligned}
 Y_t &= A'L^{1-\alpha-\nu}K_t^\alpha E_t^\nu = AK_t^\alpha E_t^\nu \\
 E_t &= F_t + BR_t^\gamma \\
 R_{t+1} &= R_t(1 - \delta) + I_t \\
 K_{t+1} &= K_t(1 - \delta) + AK_t^\alpha(F_t + BR_t^\gamma)^\nu - C_t - I_t \\
 S_{t+1} &= S_t - F_t \\
 F_t &\leq S_t \quad , \quad \forall t
 \end{aligned}$$

Preferences - use log utility, so that the value function:

$$V_t(K_t, R_t, S_t) = \log C_t + \beta V_{t+1}(K_{t+1}, R_{t+1}, S_{t+1})$$

First Order Conditions:

$$\begin{aligned}
\text{w.r.t. } C_t, \quad \beta \frac{\partial V_{t+1}}{\partial K_{t+1}} &= \frac{1}{C_t} \\
\text{w.r.t. } I_t, \quad \frac{\partial V_{t+1}}{\partial K_{t+1}} &= \frac{\partial V_{t+1}}{\partial R_{t+1}} \\
\text{w.r.t. } F_t, \quad \mu_t &= \beta \nu A K_t^\alpha (F_t + B R_t^\gamma)^{\nu-1} \frac{\partial V_{t+1}}{\partial K_{t+1}} - \beta \frac{\partial V_{t+1}}{\partial S_{t+1}}
\end{aligned}$$

Envelope Theorem Conditions:

$$\begin{aligned}
\text{w.r.t. } K_t, \quad \frac{\partial V_t}{\partial K_t} &= \beta(1 + \alpha A K_t^{\alpha-1} (F_t + B R_t^\gamma)^\nu - \delta) \frac{\partial V_{t+1}}{\partial K_{t+1}} \\
\text{w.r.t. } R_t, \quad \frac{\partial V_t}{\partial R_t} &= \beta \nu \gamma A B R_t^{\gamma-1} K_t^\alpha (F_t + B R_t^\gamma)^{\nu-1} \frac{\partial V_{t+1}}{\partial K_{t+1}} + \beta(1 - \delta) \frac{\partial V_{t+1}}{\partial R_{t+1}} \\
\text{w.r.t. } S_t, \quad \frac{\partial V_t}{\partial S_t} &= \beta \frac{\partial V_{t+1}}{\partial S_{t+1}} + \max(\mu_t, 0)
\end{aligned}$$

such that:

$$\begin{aligned}
F_t < S_t &\Rightarrow \mu_t = 0 \\
F_t = S_t &\Rightarrow \mu_t > 0
\end{aligned}$$

Eliminate marginal values to derive difference equations to characterise the system. 6 equations in 6 unknowns, $K_t, R_t, S_t, C_t, F_t, \mu_t$. We know K_0, R_0, S_0 . We shall need to construct an algorithm to determine $C_0, F_0 \leq S_0$. Then (assuming $F_0 < S_0$) for times $1 \leq t < T$:

$$\begin{aligned}
S_t &= S_{t-1} - F_{t-1} \\
K_t + R_t &= (K_{t-1} + R_{t-1})(1 - \delta) + A K_{t-1}^\alpha (F_{t-1} + B R_{t-1}^\gamma)^\nu - C_{t-1} \\
C_t &= C_{t-1} \beta (1 + \alpha A K_t^{\alpha-1} (F_t + B R_t^\gamma)^\nu - \delta) \\
\frac{\alpha}{K_t} &= \frac{\nu \gamma B R_t^{\gamma-1}}{F_t + B R_t^\gamma} \\
\mu_t &= 0 \\
\left(\frac{K_t}{K_{t-1}}\right)^\alpha \left(\frac{F_t + B R_t^\gamma}{F_{t-1} + B R_{t-1}^\gamma}\right)^{\nu-1} &= 1 + \alpha A K_t^{\alpha-1} (F_t + B R_t^\gamma)^\nu - \delta
\end{aligned}$$

Time T is defined as the point, t , at which the first system above generates $F_t > S_t$. From $t \geq T$ the system is specified by:

$$\begin{aligned}
F_t &= S_t \\
S_t &= S_{t-1} - F_{t-1} \\
K_t + R_t &= (K_{t-1} + R_{t-1})(1 - \delta) + AK_{t-1}^\alpha (F_{t-1} + BR_{t-1}^\gamma)^\nu - C_{t-1} \\
C_t &= C_{t-1} \beta (1 + \alpha AK_t^{\alpha-1} (F_t + BR_t^\gamma)^\nu - \delta) \\
\frac{\alpha}{K_t} &= \frac{\nu \gamma BR_t^{\gamma-1}}{F_t + BR_t^\gamma} \\
\mu_t &= \nu A \left(K_t^\alpha (F_t + BR_t^\gamma)^{\nu-1} \frac{1}{C_t} - \beta K_{t+1}^\alpha (F_{t+1} + BR_{t+1}^\gamma)^{\nu-1} \frac{1}{C_{t+1}} \right)
\end{aligned}$$

The time t value of each asset K_t, R_t, S_t is:

$$\begin{aligned}
\frac{\partial V_t}{\partial K_t} &= (1 + \alpha AK_t^{\alpha-1} (F_t + BR_t^\gamma)^\nu - \delta) \frac{1}{C_t} \\
\frac{\partial V_t}{\partial R_t} &= (1 + \nu \gamma A BR_t^{\gamma-1} K_t^\alpha (F_t + BR_t^\gamma)^{\nu-1} - \delta) \frac{1}{C_t} \\
\frac{\partial V_t}{\partial S_t} &= \nu AK_t^\alpha (F_t + BR_t^\gamma)^{\nu-1} \frac{1}{C_t}
\end{aligned}$$

Therefore we calculate the total return on each asset over period $(t, t + 1)$. By a no arbitrage argument these total returns should all be the same, and since there is no uncertainty, their value should be $1/\beta$.

$$\begin{aligned}
TR(K_t) &= \left[(1 - \delta) \frac{\partial V_{t+1}}{\partial K_{t+1}} + \frac{\alpha AK_t^{\alpha-1} (F_t + BR_t^\gamma)^\nu}{\beta C_t} \right] / \frac{\partial V_t}{\partial K_t} \\
TR(R_t) &= \left[(1 - \delta) \frac{\partial V_{t+1}}{\partial R_{t+1}} + \frac{\nu \gamma A BR_t^{\gamma-1} K_t^\alpha (F_t + BR_t^\gamma)^{\nu-1}}{\beta C_t} \right] / \frac{\partial V_t}{\partial R_t} \\
TR(S_t) &= \frac{\partial V_{t+1}}{\partial S_{t+1}} / \frac{\partial V_t}{\partial S_t}
\end{aligned}$$

Steady state can be calculated:

$$\begin{aligned}
S^* &= 0 \\
F^* &= 0 \\
K^* &= \left(\frac{\alpha AB^\nu (\gamma\nu/\alpha)^{\gamma\nu}}{1/\beta + \delta - 1} \right)^{\frac{1}{1-\alpha-\gamma\nu}} \\
R^* &= \frac{\gamma\nu}{\alpha} K^* \\
C^* &= AB^\nu (K^*)^\alpha (R^*)^{\gamma\nu} - \delta(K^* + R^*) \\
\mu^* &= \frac{\nu AB^{\nu-1} (K^*)^\alpha (R^*)^{\gamma(\nu-1)} (1-\beta)}{C^*} \\
E^* &= B(R^*)^\gamma
\end{aligned}$$

The algorithm to determine $C_0, F_0 \leq S_0$

- Given F_0 , adjust C_0 so that C_t is arbitrarily close to C^* at some suitably large t (i.e use the *forward shooting* method)
- If the total return on fossil fuels over the final period of their use is too high then they are in too short supply and should be conserved at the outset i.e. F_0 is too high and should be lowered. Vice versa for total returns that are too low.

Calibration details:

- Assume the calibration of Golosov, Hassler, Krusell, and Tsyvinski (2011) i.e. $\alpha = 0.3$, $\nu = 0.03$, $\beta = 0.985$ (per annum), $\delta = 1$. Time step = 10 years. Therefore still have 3 unknown parameters: A, B , & γ .
- Assume initial alternative energy capital stock $R_0 = 0$
- Estimate initial global capital stock, K_0^7 .
- Calibrate the TFP parameter, $A = YK_0^{-\alpha}E^{-\nu}$ by taking 2011 GWP from the data and estimating E by taking the percentage of global primary energy supply from fossil fuels from data ($\sim 80\%$) and combining with initial carbon usage from the laissez-faire version of Golosov, Hassler, Krusell, and Tsyvinski (2011), $128GtC$, so that $E = 128/80\%$.
- Following Golosov, Hassler, Krusell, and Tsyvinski (2011) let $S_0 = 1400$ in laissez-faire ($400GtC$ from oil and $5000GtC$ from coal which has an efficiency of 0.2 and convert to energy units (1GtC of oil gives 1 energy unit) rather than units of carbon). The Golosov,

⁷1990 estimate of global capital stock taken from Nehru and Dhareshwar (1993), rolled up to 2011 using Gross World Product from Wikipedia

Hassler, Krusell, and Tsyvinski (2011) *optimum* uses 691 energy units of fossil fuel and is implemented using an initial carbon tax of $\$56.9/tC$.

- Calibrate B & γ , using $S_0 = 691$ (i.e. the Golosov, Hassler, Krusell, and Tsyvinski (2011) *optimum*), by
 1. assuming that $E_0 = 2E^*$ (i.e. we assume that a future world which uses non-fossil energy resources will be more energy constrained than our current world - this may or may not be a reasonable assumption, and it is important for the timepath of carbon taxes.
 2. matching the initial tax rate of $\$56.9/tC$.
- The policy we cost here, from Allen, Frame, Huntingford, Jones, Lowe, Meinshausen, and Meinshausen (2009), is restricting S_0 to 420 energy units (400GtC of oil and 100GtC from coal at 0.2 efficiency, to get the 1TtC cumulative emissions)⁸. This model run is labelled as “500GtC” in Figures 5.1, 5.2 & 5.3.
- The required carbon tax in the 500GtC policy scenario is found to be $\$70.5/tC$ at the outset.

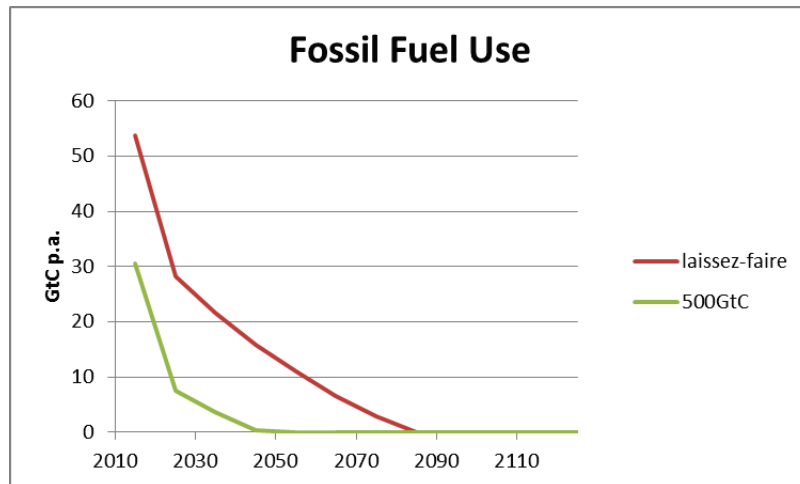


Figure B.1: Fossil fuel use, in energy units such that 1EU is provided by 1GtC of oil or 5GtC of coal, vs time. The caption on Figure 5.1 in the main body of the text does not describe this complication.

⁸The model actually produces fossil energy taxes applied to energy units and does not differentiate between oil and coal. This is not too inappropriate for the 500GtC run since coal is not much used, it is less appropriate the more coal that is in the mix.

The interaction of scale economies and energy quality

Abstract

The defining feature of the natural resources literature is that resource scarcity is associated with high resource prices that incentivise the exploitation of marginal resources and the usage of alternatives. In this paper, this incentive is higher profitability in the extractive sector, rather than simply a higher price for its output. Since energy is an essential input to the economy, its supply affects the marginal products of other factors, and in general equilibrium its supply affects the costs that the extractive industry faces. Energy sector profitability is therefore ambiguously affected by the quality of available energy resources. There are conditions related to the returns to scale in the economy which can cause lower energy sector profitability with lower energy quality. This means that marginal resources may be abandoned as high quality resources are lost. An economy which exhibits constant, or weakly increasing, returns to scale can operate at any level of energy quality, since profitability rises with falling energy quality and we observe results consistent with the usual Hotellings Rule. However an economy which exhibits strongly increasing returns to scale cannot operate with only low quality energy resources and profitability may fall with falling energy quality. It is therefore possible that an energy quality shock disincentivises, rather than incentivises, the use of marginal resources and alternatives. Ultimately, a strongly increasing returns to scale economy may have no steady state equilibrium under a decentralised market allocation, despite such an allocation being technologically feasible.

1 Introduction

The purpose of this paper is to explore the economy's response to being faced with only lower quality energy resources, with a view to characterising the situations under which this is problematic. By "Problematic" in this context, I mean that we are unable to use all available resources; or that as resources become scarcer, we receive a price signal that makes the use of resources less efficient, and does not incentivise the use of alternatives. The scenario considered is that capital assets are used to supply energy to the economy, but the manufacture of capital assets can be an energy intensive process. If higher quality energy resources are no longer available, and the use of lower quality resources is to be expanded by applying more capital inputs to their exploitation, then the relative energy (output) to capital (input) price movement will have to be consistent with this expansion. A three good economy is therefore of the minimum complexity required to investigate this issue, and given the feedbacks between available energy resources and economy wide prices, a general equilibrium approach is appropriate.

In the theoretical natural resource economics literature, natural resource scarcity is accompanied by a rise in their price. This leads to fairly sanguine conclusions with regards to the exhaustion of non-renewable resources. Since [Hotelling \(1931\)](#), the defining characteristic of the optimal depletion of non-renewable resources is that resource prices should rise at a rate related to the rate that can be earned by extraction and investing the financial proceeds. This rising price ensures that resources that are initially unprofitable to exploit eventually become profitable, and that there are incentives both to economise on the use of the resources and to develop alternatives. This foundation to the literature has subsequently been built upon, e.g. [Holland \(2008\)](#) describes models of resource extraction that generate a peak in the extraction rate during the extraction period using a partial equilibrium approach (since interest rates and backstop prices do not depend upon energy used in the aggregate economy); [Dasgupta and Heal \(1974\)](#) extend The Hotelling's framework to a general equilibrium setting without changing the conclusion that non-renewable resource prices rise without bound as they become more scarce; and [Aghion, Howitt, Brant-Collett, and Garcaia-Pealosa \(1998\)](#) describe a two sector general equilibrium model in which growth can be sustained despite declining availability of non-renewable natural resources, that are essential for production, through investment in intellectual capital. In all these cases, natural resource scarcity is accompanied by a rise in their price. Holland claims that price movements will be smoothly increasing because "oil is virtually costless to store in its natural reservoir ... even completely myopic firms without secure property rights would wait to produce from these [higher cost] deposits until the price were high enough to cover the extraction costs". This statement reveals, I believe, a possible shortcoming in this approach: yes, the resources can be left in the ground at zero cost, but there is no guarantee that the intermediate goods which are used to extract these resources

will be reasonably priced in future. It may be the case that as resources become scarce, the price of intermediates rises faster than energy prices, and so lower quality energy resources can never be profitably exploited.

The model presented in this paper has several distinctive features. A multi-sector economy is necessary in order for endogenous energy sector input prices, and heterogeneous energy resource quality is necessary since we need a marginal firm who decides to produce or to exit. However, the analysis reveals that economies of scale in the intermediate goods sector play a crucial role in determining how the economy responds to declining availability of energy resources: Constant returns, or a low level of increasing returns, are consistent with energy scarcity causing energy prices to rise faster than the prices of intermediates, and so for the economy to profitably expand into lower quality energy resources; However, if the degree of returns to scale in the intermediate goods sector is strong enough then the reduction in this sector's productivity, caused by the restriction in its factor inputs, boosts the price of intermediates by more than the price of energy. The supply of energy therefore contracts rather than expands at the margin, and eventually the economy can collapse.

Models with increasing returns are widespread in other fields e.g. economic geography models with agglomeration effects; business cycle models with increasing returns as a partial explanation for size of fluctuations; endogenous growth models; and new-trade models of intra-industry trade. The mechanism used in this paper and which is common to many papers in the literature is well described in [Ventura \(2005\)](#)¹ in which the cost of final goods production is falling in the number of inputs (increasing returns to scale), but the number of inputs depends on demand from producers of final goods. In [Ventura \(2005\)](#), this leads to a multiplicity of equilibrium locations chosen by industrial sectors. In this paper we see the same non-linear effect as the cost of energy production is falling in the productivity of intermediate inputs, but the productivity of intermediate input production depends on demand from energy producers (increasing returns to scale), which is partially determined by the quality of available energy resources. This interaction between possible increasing returns to scale and energy resource limitations has not previously been considered, and it is easy to imagine that it may be important. For example perhaps the ability and profitability of deep oil drilling is only possible because there is a full manufacturing supply chain that is predicated on the existence of automobile and aerospace industries. Perhaps if there was no cheap oil available, the contraction of the automobile and aerospace sectors could affect the manufacturing supply chain in such a way as to drastically increase costs/decrease productivity in the sector that manufactures equipment for deep oil drilling. This in turn may mean that, without cheap oil sustaining the automobile and aerospace sectors, the deep water drilling sector is unprofitable, and so it would not exist in cheap oil's absence.

¹Section 3.2 of [Ventura \(2005\)](#).

Given that a large intermediate sector lowers the cost of energy production, whilst a high level of energy production enables a large intermediate sector; the model here presents the possibility for development to go into reverse as this dependence, combined with the loss of energy resources, causes this productive equilibrium to be destabilised. This is suggestive of “The Big Push” story of economic development of [Murphy, Schleifer, and Vishny \(1989\)](#), in which coordinated industrial investment can make such investment profitable (when an investment at the margin would have been unprofitable) through the impact that the coordinated investment has on scale and efficiency. However, in [Murphy, Schleifer, and Vishny \(1989\)](#), endowments are constant and development is a coordination problem; whereas here I explicitly consider a decline (through resource depletion) in the endowment, such that the industrial equilibrium is destabilised² i.e. the issue here is the very existence of equilibrium rather than the need to coordinate on a better equilibrium.

The contribution that this article makes is to draw attention to the possibility that price movements in response to scarcity may not be favourable to bringing on substitutes or for using capital intensive but energy efficient alternatives, a possibility which is not considered in the existing literature. The paper is structured as follows: section 2 introduces the macroeconomy and section 3 presents some propositions which capture the mechanism underlying the basic argument presented. The energy sector in section 2 & 3 is completely abstract and has a very simple characterisation. Section 4 presents a specific ‘*Renewable Energy*’ model which satisfies the characterisation of the energy sector given in section 2. Section 5 does likewise with a ‘*Fossil Fuel*’ sector. Section 6 presents some illustrative results from the model, section 7 discusses the interaction between conclusions from this model and the incentives to innovate with policy implications, and section 8 concludes.

2 The macroeconomy

Households and the final goods sector in this model are standard. Final goods are produced using a constant returns, Cobb-Douglas technology, under perfect competition, using energy services,

²And I further do not have the “cottage” sector of [Murphy, Schleifer, and Vishny \(1989\)](#) to sustain output if there is no industrial sector.

intermediate goods, and labour as factor inputs. We can write:

$$Y_t = a_t L_t^{1-\mu-\gamma} E_t^\mu Q_t^\gamma = C_t + \dot{K}_t + \delta K_t = w_t L_t + p_K(t) K_t + \Pi_t$$

where

- $Y_t \equiv$ output flow at time t
- $a_t \equiv$ TFP at time t
- $L_t \equiv$ Labour inputs at time t
- $E_t \equiv$ Energy inputs at time t
- $Q_t \equiv$ Intermediate good inputs at time t
- $C_t \equiv$ Consumption at time t
- $K_t \equiv$ Capital stock at time t
- $w_t \equiv$ Wage rate at time t
- $p_K(t) \equiv$ Rental rate of capital at time t
- $\Pi_t \equiv$ Energy sector profit flow rate at time t
- $\mu \equiv$ Income share paid to energy
- $\gamma \equiv$ Income share paid to intermediates
- $\delta \equiv$ Capital depreciation rate

For simplicity assume that TFP is constant and that population growth is zero, so we can write $Y_t = A E_t^\mu Q_t^\gamma$ (where $A \equiv a_t L_t^{1-\mu-\gamma} = \text{const}$). Households have CRRA preferences, and maximise lifetime utility taking the paths of wages, rental rates of capital and energy sector profits as given. Therefore the standard consumption Euler equation holds:

$$\frac{\dot{C}_t}{C_t} = \frac{p_K(t) - \delta - \rho}{\epsilon}$$

where

- $\rho \equiv$ rate of time preference
- $\epsilon \equiv$ coefficient of relative risk aversion

An interest rate, $p_K(t)$, less (greater) than $\rho + \delta$ implies that consumption is falling (rising), and consumption, C_t , greater (less) than $A E_t^\mu Q_t^\gamma - \delta K_t$ implies that capital stock is falling (rising).

Constant returns and perfect competition in the final goods sector give us the prices of energy

and intermediate goods:

$$\begin{aligned}
p_E(t) &= \mu Y_t / E_t \\
p_Q(t) &= \gamma Y_t / Q_t \\
\text{define } z_t &\equiv \frac{p_E(t)}{p_Q(t)} \\
\Rightarrow Q_t &= \frac{\gamma}{\mu} z_t E_t \\
\text{so } p_Q(t) &= A \gamma^\gamma \mu^{1-\gamma} z_t^{\gamma-1} E_t^{\mu+\gamma-1}
\end{aligned}$$

There are two further sectors in this economy: an energy sector which uses intermediate goods to access energy resources and supply energy services; and an intermediate goods sector which rents the capital stock and produces intermediate goods. The properties that we place on the energy sector in this section and the next are that:

$$\begin{aligned}
E_t &= f(Q_E(t)) = E(z_t, s_t) \geq 0 \\
Q_E(t) &= Q_E(z_t, s_t) \geq 0 \\
\text{where } & Q_E(t) \text{ is the quantity of intermediate goods used} \\
& f \text{ is the energy from intermediates production function} \\
& z_t \text{ is the energy to intermediates price ratio, taken as given} \\
& s_t \text{ is an index, increasing in energy resource quality} \\
\text{s.t. } & \frac{\partial E_t}{\partial z_t}, \frac{\partial Q_E(t)}{\partial z_t} > 0 \\
\exists z_{min}(s_t) & \text{ s.t. } E(z_{min}(s_t), s_t) = Q_E(z_{min}(s_t), s_t) = 0 \\
& \frac{dz_{min}(s_t)}{ds_t} < 0, \text{ \& } z_{min} \rightarrow \infty \text{ as } s \rightarrow 0 \\
\lim_{z \rightarrow \infty} E_t &= a(s_t) z_t^x, \quad x \geq 0 \\
\lim_{z \rightarrow \infty} \left[\frac{Q_E(t)}{E_t} \right] &= b(s_t) z_t^y, \quad 0 \leq y \leq 1 \\
\lim_{z \rightarrow z_{min}} E_t &= c(s_t) g(z_t), \quad g(z_{min}) = 0 \\
\lim_{z \rightarrow z_{min}} Q_E(t) &= d(s_t) g(z_t)
\end{aligned}$$

This is a fairly general specification for an energy sector, with output (and demand for inputs) increasing in the relative price of energy output to intermediates input, energy output increasing in the energy quality index, and with some simple regularity assumptions on the limiting behaviour at high and low price ratios. This general specification is sufficient to generate the phenomena described in this article, but two specific, not so general, details are required in order to observe

these results. The first feature is price taking behaviour. This model assumes infinitesimal profit maximising firms who take prices as given. In these circumstances, the firms cannot internalise the effects of their supply decisions on economy wide prices. This is crucial and trivially important when elucidating a flaw in the allocation that may arise from a decentralised market: we would see nothing interesting if we analysed the social planner's solution (as was done in e.g. [Dasgupta and Heal \(1974\)](#) in their general equilibrium Hotellings model) or if we analysed the solution with a monopoly energy supplier.

The other important specific feature that this energy sector must exhibit is a minimum price for its output relative to its inputs at which it is willing to produce a positive quantity. The specific combination of price taking (which implies infinitesimal firms and decreasing returns to scale) and a minimum price at which these firms are willing to operate, calls for fixed costs in order to prevent productivity rising without bound as the scale of production goes to zero. This seems reasonable in context since it would seem that fixed costs are a realistic feature in the energy industry: the output of oil from the application of a very small quantity of deep water drilling equipment is not high, with decreasing returns to additional units of equipment, rather it is likely to be zero because these additional units of equipment are essential; likewise the output from a wind or solar farm that is disconnected from the grid is zero or very small, and the output from a wind turbine blade or silicon wafer in the absence of the rest of the components is definitely zero.

Finally it is important to note that whilst these features for the energy industry are necessary in order to see the phenomena described in this paper, the phenomena are not a necessary consequence of these features. It is not the case that describing a price taking energy industry with fixed costs is equivalent to assuming the results presented. We shall see that there is a large parameter region in which results are *standard*, and the interesting *new results* are only exhibited in the presence of a sufficient degree of scale economies in the intermediate goods sector.

The intermediate goods sector is zero profit making, and rents capital stock to create intermediate goods. We write the intermediate goods production function as:

$$Q_t + Q_E(t) = \theta K_t^\psi, \quad \psi \geq 1 \quad (1)$$

where ψ determines the degree of returns to scale. θ is a normalisation parameter. Zero profits implies:

$$\begin{aligned} p_Q(t)(Q_t + Q_E(t)) &= p_K(t)K_t \\ \text{i.e. } p_K(t) &= \theta^{\frac{1}{\psi}} p_Q(t)(Q_t + Q_E(t))^{\frac{\psi-1}{\psi}} \end{aligned} \quad (2)$$

There is some evidence to suggest that manufacturing industries behave as if they are subject

to increasing returns to scale. [Hall \(1989\)](#) explains the correlation of factor productivity with exogenous demand shocks using increasing returns and finds that increasing returns are particularly evident in the aggregate economy and in manufacturing sectors. [Caballero and Lyons \(1989\)](#) split the returns to scale evident in the aggregate economy into internal, firm level, constant or decreasing returns to scale, and positive external returns to scale. Their best estimate of the degree of scale economies in the US is that a sector which increases its inputs by 10% will see an increase in output of 8%, but if the whole economy increases its inputs by 10% then output will rise by 13%³. [Basu and Fernald \(1997\)](#) explain similar data as [Caballero and Lyons \(1989\)](#) as a reallocation effect towards more efficient firms rather than any real increasing returns at the micro-level, but agree that if we model the aggregate economy as a representative firm then increasing returns to scale are appropriate.

To simplify the analysis I consider only steady states of this economy. The consumption Euler equation implies that in steady state the rental rate of capital is a constant given by:

$$p_K^* = \rho + \delta \quad (3)$$

If rental rates are below this steady state rate, then households will be reducing their holdings of capital by saving at a rate that implies overall capital stocks are falling. Whilst if rental rates are above this steady state rate, then households will be increasing their holdings of capital by saving at a rate that implies overall capital stocks are rising. Equation (2) can be used to give a further condition on the rental rate of capital as a function of the energy to intermediates price ratio that can pin down the steady state of the whole economy, where we now drop the time subscripts to indicate that we consider only steady states of the economy:

$$p_K(z, s) = \theta^{\frac{1}{\psi}} A \gamma^\gamma \mu^{1-\gamma} z^{\gamma-1} E(z, s)^{\mu+\gamma-1} \left(\frac{\gamma}{\mu} z E(z, s) + Q_E(z, s) \right)^{\frac{\psi-1}{\psi}} \quad (4)$$

Clearly we always have $p_K(z, s) \geq 0$, $\forall z \in [z_{min}, \infty)$ and, by monotonicity of $E(z, s)$ and $Q_E(z, s)$ with respect to z , $p_K(z, s)$ is continuous over this set.

³i.e. in the notation of equation (1), this translates as $\psi \sim 1.3$ (ignoring the fact that capital services are only a subset of the whole economy, which is also subject to diseconomies of scale caused by declining energy resource quality). In the empirical trade exercise of [Mohler and Seitz \(2010\)](#), the elasticity of substitution in the CES import demand systems of European economies is found to lie in the range 3 - 5, which corresponds to a returns to scale parameter, $\psi \in (1.25, 1.5)$

3 Steady State Equilibrium

Given some energy quality s , the intersections of the equation (4) with the constant steady state value given by equation (3) defines steady state values, z^* . In this section I show that the parameter space for this economy can be divided into three: one parameter region that correspond to the common understanding of how prices respond to scarcity; one parameter region that can be dismissed as unrealistic, and one interesting *new region* that is the contribution of this paper.

Proposition 1. *Trivially, $K = 0$ is a steady state since, given zero capital stock, production and so investment is zero.*

Proposition 2. *For any given capital stock, $K > 0$, the market equilibrium exists and is unique.*

The equilibrium price ratio is related to capital stock by:

$$K = \left(\frac{Q + Q_E}{\theta} \right)^{\frac{1}{\psi}} = \left(\frac{\frac{\gamma}{\mu} z E(z, s) + Q_E(z, s)}{\theta} \right)^{\frac{1}{\psi}}$$

i.e. $K'(z) > 0$, since $\frac{\partial E}{\partial z}, \frac{\partial Q_E}{\partial z} > 0$

$z \in (z_{min}(s), \infty)$ represents all possible price ratios that are associated with positive output from the energy sector. Clearly $K(z)$ is a bijection on $z > z_{min}(s)$ and so a given capital stock, K , will imply a particular price, z , by equation (4)⁴. There are therefore no problems of interpretation with a multiplicity of equilibria (though as we shall see, there may be multiple steady states). This monotone relationship between K and z , as well as the relationship already derived, Equation 4, between p_K and z , allows us to construct phase diagrams in (K, C) space based on the consumption Euler equation and the equation of motion for capital. First however, we need to characterise how Equation 4 behaves for different values of ψ .

Proposition 3. $\exists \psi^{**} > 1$ such that $\psi > \psi^{**} \Rightarrow p_K(z, s) \rightarrow \infty$ as $z \rightarrow \infty$ and $\psi < \psi^{**} \Rightarrow p_K(z, s) \rightarrow 0$ as $z \rightarrow \infty$.

⁴Note that this is a statement for all K , not just at the steady state K implied by the parameters of the model and by the energy quality index s . The renewables energy sector of section 4 has no energy sector dynamics and so this proposition holds whether or not the macroeconomy has settled into its steady state. However, the fossil fuel energy sector of section 5 has only had its price quantity relationships analysed assuming that it faces constant (i.e. steady state) prices. Therefore the signs of the partial derivatives of E and Q_E for this sector, although they are positive with respect to z as required, have strictly only been evaluated at steady state.

We can evaluate ψ^{**} by taking limits of equation (4):

$$\begin{aligned} \text{Have } p_K(z, s) &= A\theta^{\frac{1}{\psi}}\gamma^\gamma\mu^{1-\gamma}z^{\gamma-1}E(z, s)^{\mu+\gamma-1}\left(\frac{\gamma}{\mu}zE(z, s) + Q_E(z, s)\right)^{\frac{\psi-1}{\psi}} \\ \text{so } \lim_{z \rightarrow \infty} [p_K(z, s)] &= \text{const}_1 \times \lim_{z \rightarrow \infty} \left[z^{\gamma-1}E(z, s)^{\mu+\gamma-1+\frac{\psi-1}{\psi}} \left(\frac{\gamma}{\mu}z + \frac{Q_E(t)}{E_t}\right)^{\frac{\psi-1}{\psi}} \right] \\ &= \text{const}_2 \times \lim_{z \rightarrow \infty} \left[z^{\gamma-1+x(\mu+\gamma-1+\frac{\psi-1}{\psi})+\frac{\psi-1}{\psi}} \right] \end{aligned}$$

$$\begin{aligned} \text{Therefore } p_K(z, s) \rightarrow \begin{matrix} \infty \\ 0 \end{matrix} \text{ as } z \rightarrow \infty &\iff \gamma - 1 + x \left(\mu + \gamma - 1 + \frac{\psi - 1}{\psi} \right) + \frac{\psi - 1}{\psi} \gtrless 0 \\ &\iff \psi \gtrless \psi^{**} = \frac{1}{\gamma + \frac{x}{1+x}\mu} \end{aligned}$$

Proposition 4. $\exists \psi^* \in (1, \psi^{**})$ such that $\psi > \psi^* \Rightarrow p_K(z, s) \rightarrow 0$ as $z \rightarrow z_{min}$ and $\psi < \psi^* \Rightarrow p_K(z, s) \rightarrow \infty$ as $z \rightarrow z_{min}$.

We can evaluate ψ^* by taking limits of equation (4):

$$\begin{aligned} \text{Have } p_K(z, s) &= A\theta^{\frac{1}{\psi}}\gamma^\gamma\mu^{1-\gamma}z^{\gamma-1}E(z, s)^{\mu+\gamma-1}\left(\frac{\gamma}{\mu}zE(z, s) + Q_E(z, s)\right)^{\frac{\psi-1}{\psi}} \\ \text{so } \lim_{z \rightarrow z_{min}} [p_K(z, s)] &= \text{const}_1 \times \lim_{z \rightarrow z_{min}} \left[g(z)^{\mu+\gamma-1+\frac{\psi-1}{\psi}} \right] \end{aligned}$$

$$\begin{aligned} \text{Therefore } p_K(z, s) \rightarrow \begin{matrix} \infty \\ 0 \end{matrix} \text{ as } z \rightarrow z_{min} &\iff \mu + \gamma - 1 + \frac{\psi - 1}{\psi} \lesseqgtr 0 \\ &\iff \psi \lesseqgtr \psi^* = \frac{1}{\gamma + \mu} < \psi^{**} \end{aligned}$$

Whilst these propositions only strictly allow us to characterise $p_K(z, s)$ as $z \rightarrow z_{min}$ or $z \rightarrow \infty$, the function would have to be very strange to have many turning points. If we make a further regularity assumption (that will be true for the results presented in section 6) that there is at most one turning point, then we can describe the economy as a function of the degree of returns to scale in the intermediate goods sector:

- $\psi > \psi^{**}$ *Super Strong Increasing Returns to Scale* (SSIRS). Since $p_K \rightarrow 0$ as $z \rightarrow z_{min}$, $p_K \rightarrow \infty$ as $z \rightarrow \infty$, and given the assumption of a maximum of one turning point, then it

must be the case that $p_K(z, s)$ is monotonically increasing in z . Therefore it will only cross the steady state value of p_K^* once (from below) at $z^* > z_{min}$. Given that $K(z)$ is a bijection, there is a single (unstable & repulsive) steady state $K^* > 0$. Equation (4) is graphed for this extreme case of SSIRS in figure 3.1. No stable productive economy (i.e. $Y^* > 0$) exists even with maximal energy resources availability. $K^* = 0$ is stable and attractive since as $K \rightarrow 0$ the rental rates paid to capital become insufficient for households to want to save enough to prevent the capital stock decaying away. Returns to scale are too strong for any this model to describe any sensible economic system and we do not consider this case further.

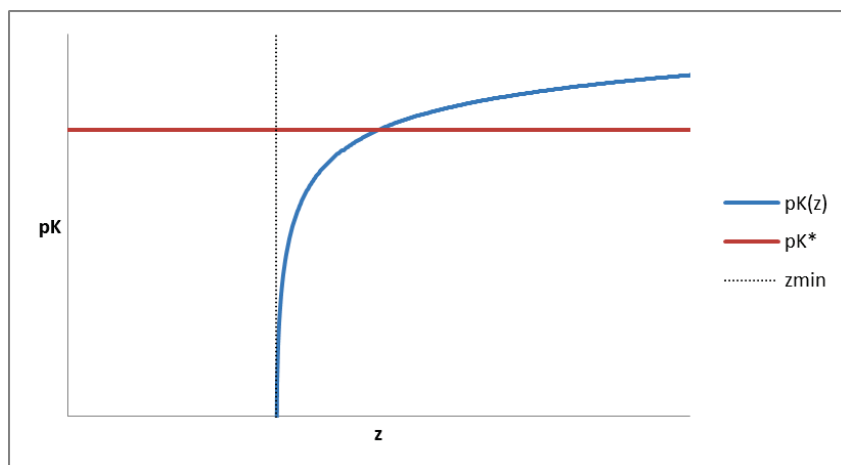


Figure 3.1: SSIRS: two steady states - a stable state at $K^* = 0$ and a higher unstable state.

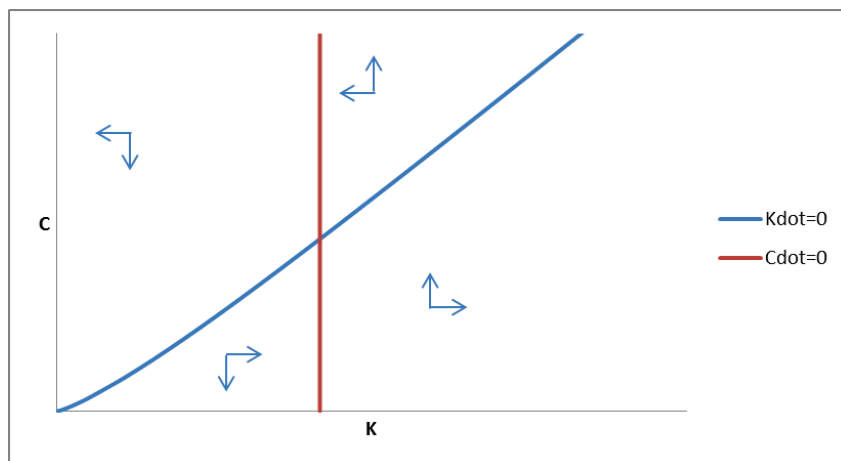


Figure 3.2: Phase diagram for SSIRS, showing the unstable steady state, $K^* > 0$, in (K, C) space

- $1 \leq \psi < \psi^*$ *Weakly Increasing (or Constant) Returns to Scale* (WIRS). Since $p_K \rightarrow \infty$ as $z \rightarrow z_{min}$ and $p_K \rightarrow 0$ as $z \rightarrow \infty$, and given the assumption of a maximum of one turning

point, then it must be the case that $p_K(z, s)$ is monotonically decreasing in z . Therefore it will only cross the steady state value of p_K^* once (from above) at $z^* > z_{min}$. Given that $K(z)$ is a bijection, there is a single (stable & attractive) steady state $K^* > 0$. Equation (4) is graphed for WIRS in figure 3.3. $K^* = 0$ is unstable and repulsive since as $K \rightarrow 0$ the rental rates paid to capital become very large and households to want to save and accumulate capital. This economy accords with our intuitions: a productive economy ($Y^* > 0$) always exists and the response of the economy to a fall in energy quality is for the value of z_{min} to rise, the whole p_K curve to shift to the right, and the equilibrium energy price to intermediates price to rise. This relative energy price rise incentivises the full usage of energy resources by endogenously bringing previously unprofitable marginal resources into use.

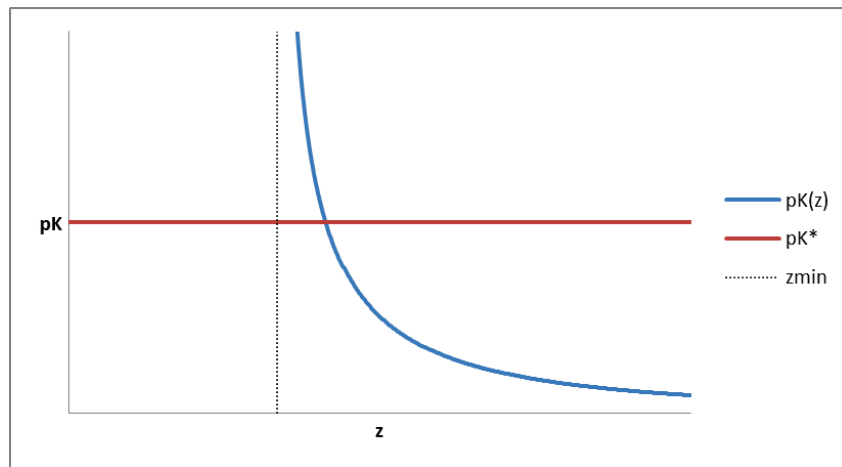


Figure 3.3: WIRS: two steady states - an unstable state at $K^* = 0$ and a stable state at $K^* > 0$.

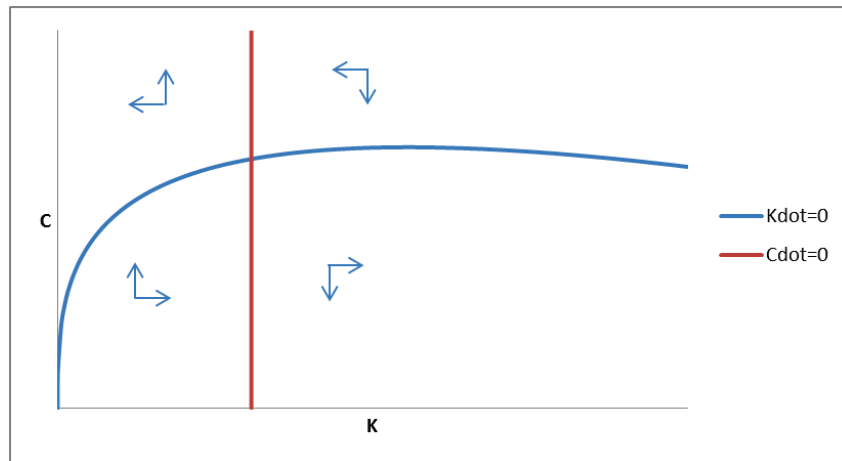


Figure 3.4: Phase diagram for WIRS, showing the stable steady state, $K^* > 0$, in (K, C) space

- $\psi^* < \psi < \psi^{**}$ *Strong Increasing Returns to Scale (SIRS)*. Now have $p_K \rightarrow 0$ as $z \rightarrow z_{min}$ and $p_K \rightarrow 0$ as $z \rightarrow \infty$, with $p_K > 0, \forall z \in (z_{min}, \infty)$. Given the assumption of a maximum of one turning point, then there must indeed be a single turning point at $z^+(s)$ given by $\partial p_K(z^+(s), s)/\partial z = 0$. $K^* = 0$ is stable and attractive since as $K \rightarrow 0$ the rental rates paid to capital become insufficient for households to want to save enough to prevent the capital stock decaying away. But, so long as $p_K(z^+(s), s) > p_{K^*}$ then there will be another stable steady state at $z^* > z^+ > z_{min}$ (corresponding to $K^* > 0$). Equation (4) is graphed for SIRS in figure 3.5, showing this stable productive steady state. At the productive equilibrium, this economy may look very similar to the WIRS economy, and may respond to a fall in energy quality in a similar way with relative price rises endogenously bringing previously unprofitable marginal resources into use. However, we cannot prove the

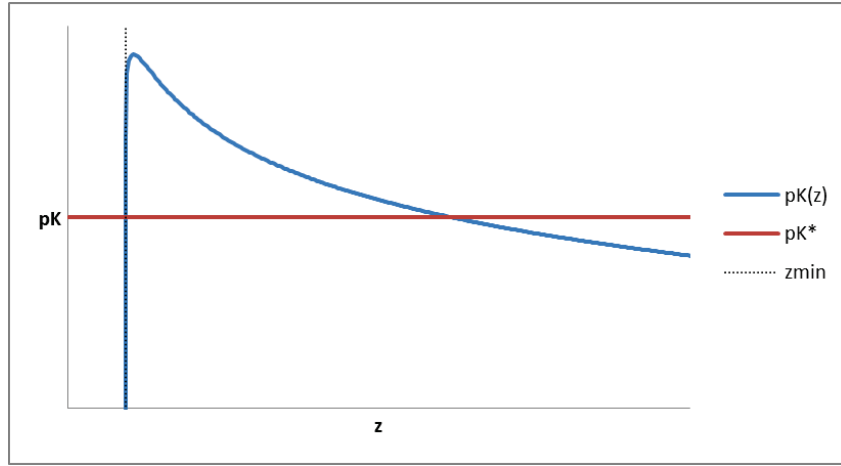


Figure 3.5: SIRS: may have three steady states: stable states at $K^* = 0$ and $K^* > 0$ with an unstable state between.

existence of a non-zero steady state for SIRS. All we know here are the limiting properties that $\lim_{z \rightarrow \infty} p_K(z, s) = \lim_{z \rightarrow z_{min}} p_K(z, s) = 0$, and that $z_{min} \rightarrow \infty$ as $s \rightarrow 0$. Therefore it is possible that the graph for Equation (4) for the SIRS economy looks like that shown in figure 3.7, i.e. with a globally stable $K^* = 0$ steady state. Indeed, we can *experimentally* construct a particular SIRS economy then with a particular $s = s_1$ such that the evaluated $p_K(z, s_1)$ function resembles figure 3.5. Then change the energy quality parameter to $s = s_2 < s_1$ such that the evaluated $p_K(z, s_2)$ function resembles figure 3.7, i.e. there is a point in the s parameter space at which the economy *collapses* as the $K^* > 0$ steady state ceases to exist. We can describe the transition from figure 3.5 to figure 3.7 as a collapse because there is a discontinuity in the steady state that the economy can reach. Once we lower the energy quality index, s , past a critical value, the steady state changes discontinuously from $K^* > 0$ to $K^* = 0$. This is unlike WIRS in which the economy exists at some positive level of

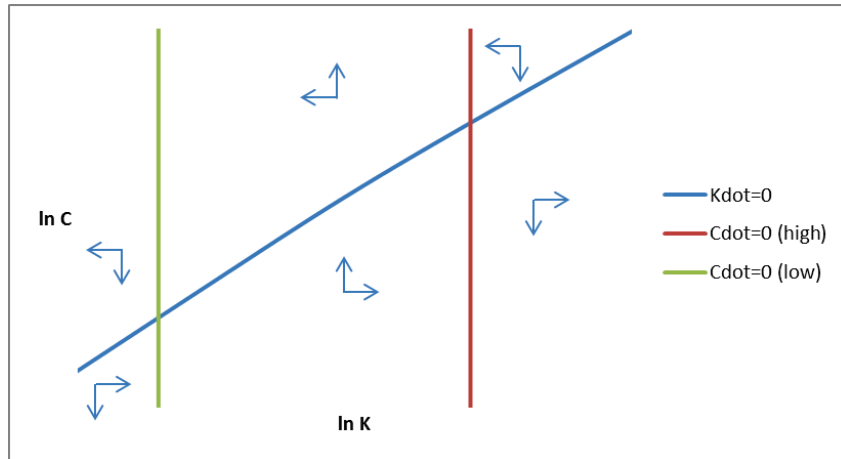


Figure 3.6: A possible phase diagram for SIRS, with 2 steady states, $K^* > 0$, the lower unstable and the higher stable, in (K, C) space. Log scale used because lower steady state close to $K = 0$.

production, irrespective of the severity of the resource restrictions, s , that are imposed.

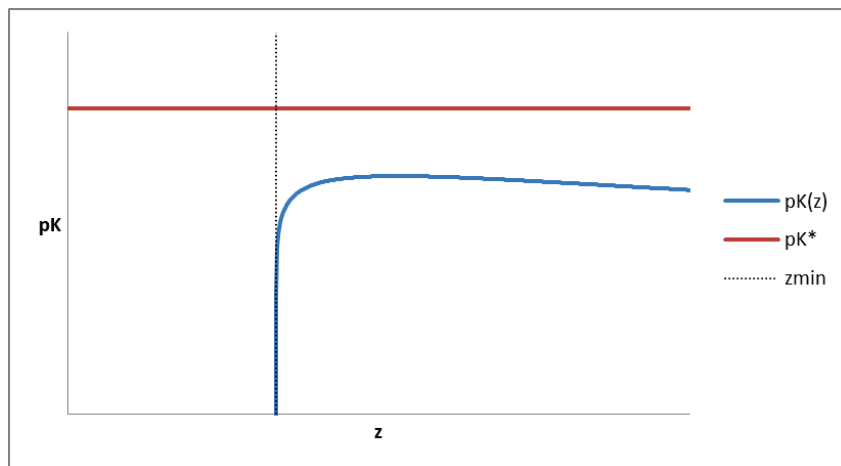


Figure 3.7: SIRS: may only have one stable steady state at $K^* = 0$.

The intuition for what is going on here is straightforward: exploiting energy resources requires intermediate goods as inputs and the scale of the energy sector will exogenously depend upon the quality of resources available, and endogenously upon the relative output to input price. An fall in energy quality is a supply shock to the energy sector which is felt throughout the whole economy. In the absence of significant scale economies in intermediate good production, this supply shock makes energy the scarce and hence expensive commodity, which mitigates the exogenous cause of the problem which was the decline in energy quality. If however scale economies are important in intermediate good production then as the economy contracts due to the effects of the exogenous decline in energy quality, productivity falls by a lot in the intermediate goods sector. This means

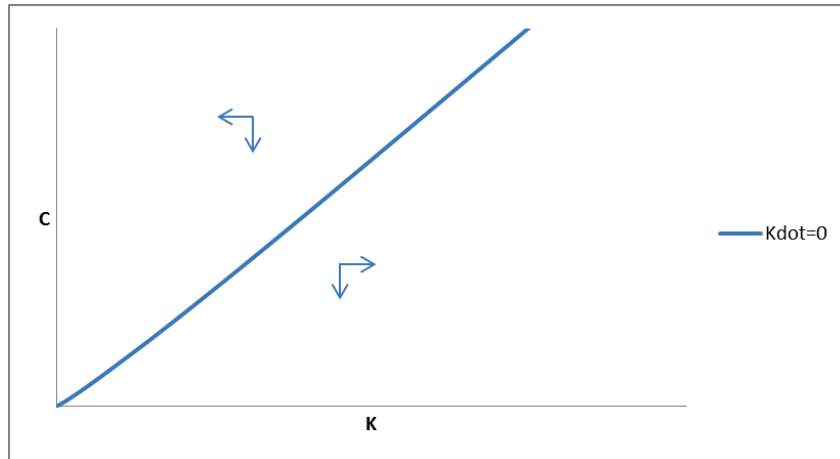


Figure 3.8: A possible phase diagram for SIRS. No steady states with $K > 0$ exists.

that intermediates become relatively scarce and expensive. The price effects move in the opposite direction needed to mitigate the exogenous cause of the problem, this exacerbates the problem by restricting energy production further. These effects multiply and eventually there is no energy price and intermediate goods price which can simultaneously produce positive output from the energy sector and allow the factors supplying the final goods sector to be paid their marginal products, whilst paying capital at the steady state interest rate. In this circumstance, even in the absence of any further declines in energy quality, the interest rate will be below the required rate of return and the economy will run down its capital stock towards the zero capital stock, zero production steady state.

4 A renewable energy sector

A model of some types of renewable energy is perfect for generating an energy sector whose behaviour is consistent with the assumptions made in section 2 and which can be studied *in steady state*. If the resources that are exploited are always there, i.e. next period's resources are not impacted by usage in this period, then there is no trade off across time that this sector needs to make. It will take prices as given now, and make an optimal choice now; the future does not matter. This describes resources like wind or solar resources available at a given site, and not timber or other biomass which needs to be managed with a more lifecycle view. Therefore in the subsequent discussion, read 'resources' perhaps as location specific wind speed or solar flux. It is appropriate to model such resources as being subject to decreasing returns to scale since there is a limit to the energy we can extract from a single location, no matter how much capital we deploy at that location. As previously discussed, it is natural that there is some fixed costs in for operating in any particular location: inada conditions leading to super-productive but miniscule factor inputs

are unrealistic. For any given set of prices, the best resources will be more profitable to exploit than more marginal resources. As the energy to intermediates price rises, more intermediates will be used exploiting a given resource, and marginal resources that previously were not exploited will now be brought into use.

Energy resources are owned by households who auction the right to exploit these resources to a continuum, $[e, \infty)$, of potential energy firms. The households therefore extract all the surplus and own the profit stream that the firms produce. The resulting energy market is competitive (i.e. price taking) with a continuum of differentiated firms, $j \in [e, \infty)$, each producing homogenous output, E_j using intermediate goods Q_j in a decreasing returns to scale production function that also exhibits costs indexed by j i.e. “high j ” firms are exploiting poorer quality energy resources than “low j ” firms and so, for a given quantity of inputs, Q_j , they produce a lower quantity, E_j of outputs. The production function is:

$$E_j = Q_j^\beta - j, \quad \beta \in (0, 1)$$

Firms maximise profits, $\pi_j = p_E E_j - p_Q Q_j = p_E (E_j - (1/z)Q_j)$, taking prices as given. This gives:

$$Q_j = (\beta z)^{\frac{1}{1-\beta}}$$

This is independent of j i.e. all energy firms use the same quantity of inputs. Therefore profits and energy output are both decreasing in j . j is endogenously defined on $[e, r]$ where $e = 1/s > 0$ is the exogenous parameter representing the highest quality energy resources available, whilst r is an endogenous variable that is defined by $\pi_r = 0$ i.e. there is free entry in the energy sector and firms continue to enter, making positive profits, until the marginal firm makes zero profits. This gives:

$$r = Q_r^\beta (1 - \beta) = (\beta z)^{\frac{\beta}{1-\beta}} (1 - \beta)$$

The total inputs and outputs from the energy sector are calculated by summing over the firms from e to r i.e.

$$\begin{aligned} E &= \int_e^r E_j dj = \frac{1}{2} (1 - \beta^2) (\beta z)^{\frac{2\beta}{1-\beta}} - e(\beta z)^{\frac{\beta}{1-\beta}} + \frac{1}{2} e^2 \\ Q_E &= \int_e^r Q_j dj = (1 - \beta) (\beta z)^{\frac{1+\beta}{1-\beta}} - e(\beta z)^{\frac{1}{1-\beta}} \\ \Pi &= \int_e^r \pi_j dj = p_E (E - \frac{1}{z} Q_E) \end{aligned}$$

The energy sector uses intermediate goods, and its output responds endogenously to the relationship between the output energy price and the input intermediates price. High quality resources

are those which require low inputs per unit of energy produced whereas low quality resources require higher inputs per unit of energy produced. There is no limit imposed upon energy availability, however these unlimited resources will be of increasingly poor quality. If it is optimal to exploit a particular resource, then it is optimal to exploit every resource of higher quality, and so the available high quality resources are always exploited. Exploitation of lower quality resources is an increasing function of the energy to intermediates price ratio.

Appendix 1 shows that this renewables energy sector satisfies the properties of the generic energy sector specified in section 2 and hence that the aggregate economy should have the steady state behaviour of section 3. Section 6 uses this renewables energy sector to generate illustrative results for the economy under regimes of constant, weakly increasing, and strongly increasing returns to scale in the intermediate goods sector.

5 A fossil fuel energy sector

In a Hotelling model of non-renewable resource extraction, the owners of the resources face a trade off between extracting and supplying these resources to market, and leaving the resources in the ground and seeing their price rise. Optimal extraction equates the value of these options, and the basic result is that as a finite resource is extracted, its price should rise to compensate those owners who do not extract immediately. This prediction of a rising price is at odds with the observed price history of non-renewable resources, and some economists e.g. [Barnett and Morse \(1963\)](#), and [Simon \(1996\)](#), have concluded that this price history is evidence of declining rather than increasing scarcity of energy resources. The explanation for this is usually technological advances. However, [Hamilton \(2011\)](#) details the history of global crude oil production over the last century and a half and finds that the production increases have been achieved mainly through the exploitation of new geographic areas, rather than predominantly through technological advances as applied to existing sources. As the scope for adding to production from new geographical areas declines, the suggestion is that the era of rising production could soon end. There are two effects going on: depletion and technological progress; and there is some dispute about which of these effects is “winning”.

A set of data that is broadly consistent with Hamilton’s interpretation is the energy return on energy invested (EROI) for fossil fuels over the past century (see figure 5.1). EROI can be considered as a *technologically adjusted* index of the cost of obtaining energy resources. So for example oil and gas from 1930 had an EROI of (greater than) 100 : 1 and so obtaining 100boe (barrels of oil equivalent) required *spending* energy (including the energy embodied in the capital used to extract the energy) that contained $\sim 1boe$ so that gross energy production would have

had to be $\sim 101boe$ to supply the final economy with this $100boe$. By 2005 oil and gas EROI was $\sim 15 : 1$ and supplying the final economy with $100boe$ would have required gross energy production of $\sim 107boe$. This increase in the cost of supplying the same amount of energy comes despite improvements in technology over the period. Extracting deep water oil in 1930 would not have cost an extra 6% over the oil that was being extracted at that time, rather it would not have been possible at all with the technology available. It is in this sense that EROI can be said to be a technologically adjusted index of the cost of obtaining these resources, and this data suggests that, even allowing for technological advances, the resources that we are extracting are becoming more costly.

Resource	Year	Magnitude (EJ/yr)	EROI (X:1)
Fossil fuels			
Oil and gas	1930	5	>100
Oil and gas	1970	28	30
Oil and gas	2005	9	11 to 18
Discoveries	1970		8
Production	1970	10	20
World oil production	1999	200	35
Imported oil	1990	20	35
Imported oil	2005	27	18
Imported oil	2007	28	12
Natural gas	2005	30	10
Coal (mine-mouth)	1950	n/a	80
Coal (mine-mouth)	2000	5	80
Bitumen from tar sands	n/a	1	2 to 4
Shale oil	n/a	0	5

Figure 5.1: Figure from [Murphy and Hall \(2010\)](#)

However, the total resource of fossil fuels is massive, though of increasingly poor quality. There are enormous quantities of low quality fossil fuel resources, like shale gas, tar sands and brown coal. Figure 5.2 shows that we are a long way from any limits in the availability of fossil fuel resources, notwithstanding any efforts on our part to leave some resources unused because of climate change concerns. Energy resources with higher costs of production (and/or low EROI) tend to be more capital intensive. As an illustrative example we can consider the wooden derricks used for Pennsylvanian oil production in the 19th century against the deep water drilling rigs used today in places like the Gulf of Mexico; or we can compare the pick and shovels used for easily accessible coal seams, to the machinery required for mountain top removal in the Appalachian Mountains. This low quality / high input requirement accords with the intuitive definition of energy quality

used in this article. Nuclear energy can also be viewed similarly with finite resources of uranium, but potentially massive resources if ‘breeder’ reactors are used to recycle the fuel. Again, breeder technology is more expensive and so can be viewed as ‘lower quality’ in this context. When combined with the possibility of technological advances that lower the cost, and effectively raise the quality, of currently unprofitable marginal resources, these non-renewable resources start to look like renewable resources, with a *regeneration rate* related to the rate of technological progress. This is clearly wrong in the limit, but it may be a good approximation to the fossil fuel energy resources over the next several centuries, and to nuclear energy resources over an even longer term.

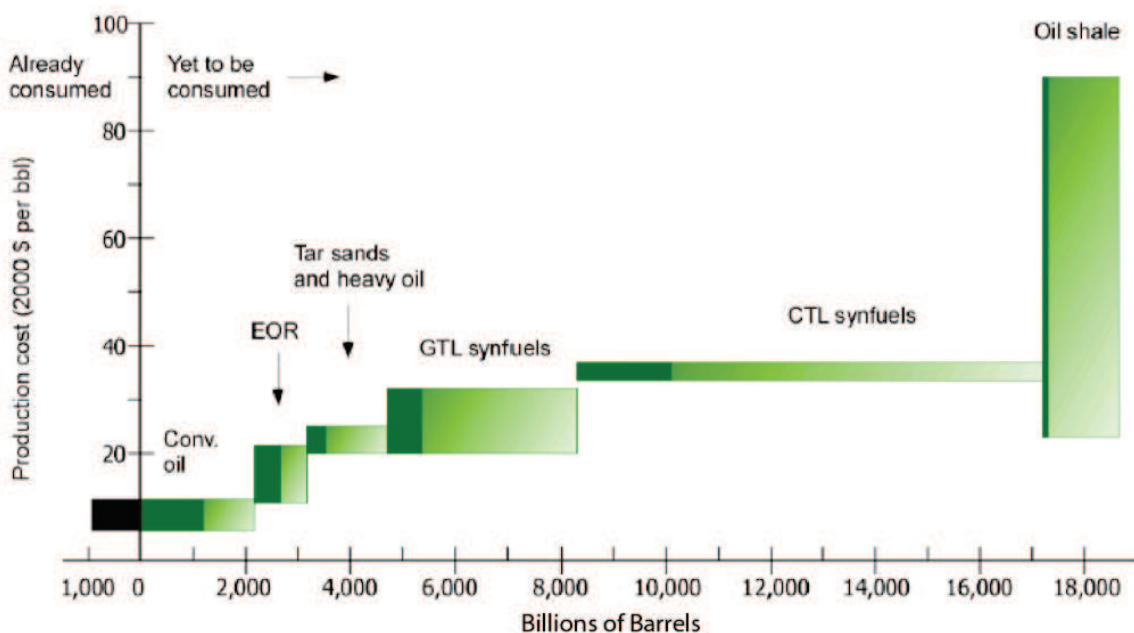


Figure 5.2: Adapted from Brandt and Farrell (2007) by Murphy (2011). Shows resources with their production cost. Proven reserves are dark bands on left, uncertain resources are lighter bands on right.

Given the expectation then (at least from some people) that technology may be sufficient to keep depletion at bay over the medium (and maybe even over the long) term, and the dispute about whether the evidence of the 20th century is consistent with this expectation, we construct a *steady state* Hotellings model of a fossil fuel energy sector in which technological progress and depletion are exactly offset, using the standard framework (see e.g. Beltratti, Chichilnisky, and Heal (1998)) for the lifecycle management of renewable resources like forests and fisheries. At steady state, these resources will again be characterised by an *energy quality* state variable, and we treat the highest quality as a parameter and endogenously allow the exploitation of lower quality resources up to a zero profit limiting case for the marginal resource. The *intellectual experiment* that is then

explored is how economies with a different level of highest energy quality available compare with each other.

Again, households own a continuum of energy firms, each indexed by j . Firms j exploits resources labelled S_j by applying inputs Q_j to produce a homogenous good E_j , taking prices as given. There are fixed costs related to the remaining size of the resource that the firm exploits (the rationale being that if high quality resources are exploited first then a large stock means there are high quality resources available, whereas a small stock implies only the poor quality resources remain). The production function and profit flows are given by:

$$\begin{aligned} E_j(t) &= (Q_j(t) - S_j(t)^{-\alpha})^\beta, \quad \beta \in (0, 1), \quad \alpha > 0 \\ \pi_j(t) &= p_E(t) \left(E_j(t) - \frac{1}{z_t} E_j(t)^{\frac{1}{\beta}} - \frac{1}{z_t} S_j(t)^{-\alpha} \right) \end{aligned}$$

In the renewable stock model, the regeneration rate of the stock is related to the current size of the stock. The simplest analytical way to avoid cornucopian solutions where the stock grows without bound is to have a quadratic equation of motion for the stock:

$$\dot{S}_j(t) = g_{j1} S_j(t) - g_2 S_j(t)^2 - E_j(t)$$

Firms maximise lifetime profits taking prices as given subject to this resource constraint. Profits are discounted at the rate of return available on capital, $r_t = p_K(t) - \delta$. The Hamiltonian of their maximisation problem is therefore:

$$H_t^c = \pi_j(t) + \lambda_t \dot{S}_j(t)$$

With solution conditions

$$\begin{aligned} \frac{\partial H_t^c}{\partial E_j(t)} = 0 &\Rightarrow \lambda_t = \frac{\partial \pi_t}{\partial E_j(t)} \equiv \pi_E(j, t) \\ \dot{\lambda}_t = r_t \lambda_t - \frac{\partial H_t^c}{\partial S_j(t)} &\Rightarrow \frac{d\pi_E(j, t)}{dt} = (r_t - g_{j1} + 2g_2 S_j(t)) \pi_E(j, t) - \pi_S(j, t) \end{aligned}$$

Without further assumptions it is difficult to go any further. If however we assume constant prices then we can describe the behaviour of firms in the fossil fuel sector by the coupled differential equations (where the time subscripts remain only because the current value of production and the

current value of the stock may be time varying, prices are constant though):

$$\begin{aligned}\dot{S}_j(t) &= g_{j1}S_j(t) - g_2S_j(t)^2 - E_j(t) \\ \dot{E}_j(t) &= \frac{1}{\pi_{EE}(j,t)} ((\rho - g_{j1} + 2g_2S_j(t))\pi_E(j,t) - \pi_S(j,t)) \\ \rho - g_{j1} + 2g_2S_j(t) &> 0 \quad (\text{for finite lifetime value})\end{aligned}$$

Substituting in for the partial derivatives of the profit function, the $\dot{S} = 0$ and $\dot{E} = 0$ loci are given by the equations:

$$\begin{aligned}\dot{S} = 0 &\Rightarrow E = g_{j1}S - g_2S^2 \\ \dot{E} = 0 &\Rightarrow E = \left(\beta z \left(1 - \frac{1}{z} \alpha S^{-\alpha-1} \frac{1}{\rho - g_{j1} + 2g_2S} \right) \right)^{\frac{\beta}{1-\beta}}\end{aligned}$$

Because the profit flow tends to minus infinity as the stock tends to zero, the lifetime profit stream associated with the path that leads to steady state is valued more highly than any paths that involve depleting the resource⁵. The system is therefore saddlepath stable, with a phase diagram of the form shown in Figure 5.3.

There can be more than one steady state for a firm in this economy. (In Figure 5.3, lowering z would lower the vertical asymptote of the $\dot{E} = 0$ locus, without changing the $\dot{S} = 0$ locus. This could lead to 3 steady states, with the actual steady state achieved being a function of the initial state.) However, this is perhaps taking the setup of this model too literally. If we were to use this literal interpretation, then we could not talk about the energy quality as a parameter: the energy quality would be an endogenous variable; the primitive parameters are g_{j1} and g_2 which describe how the resources *regenerate*. The point of this exercise is not to construct a theory of the equilibrium output of a firm supplying fossil fuels given fundamental primitives, rather it is to show that a fossil fuel sector can be broadly consistent with the properties of the abstract energy sector laid in in Section 2. Therefore, from now on we treat S_j as a *parameter* that is a property of the firm operating at this level of energy quality, and so implicitly, g_{j1} is a variable that ‘adjusts to keep S_j constant’. The actual parameter that is relevant on an economy wide basis is $s = \max(S_j : \forall j)$ and we allow firms with $S_j < s$ to enter until the marginal firm makes zero profits. We can state, by assumption, that s is such that all the firms j exploiting energy resources with quality $S_j \in (S_{min}, s)$ have a unique steady state (with a phase diagram of the form of Figure 5.3) and the position of this steady state is a smooth function of parameters and prices. Given this set up it is important to reiterate the intellectual exercise that is being undertaken: we compare

⁵Although paths with a cycle of resource overuse followed by shut-down and regeneration have not strictly been ruled out.

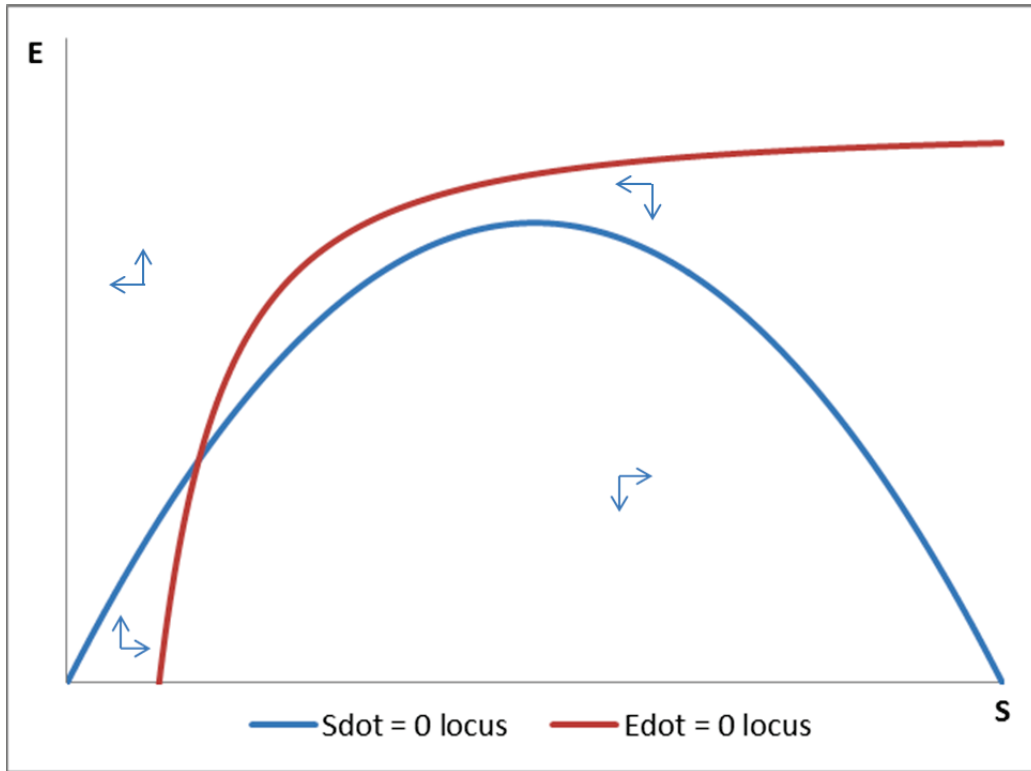


Figure 5.3: Phase diagram for fossil fuel firm, managing its resources under constant prices.

economies with different levels of s , the highest quality available energy. For $s_2 < s_1$ the discussion is framed as if s_2 is the same economy as s_1 but after some *energy quality* shock. However the firms occupying a particular *energy quality location*, S_j are not the same firms (if we assume that the primitive parameters of an individual firm are constant): conditional on g_{j1} a rise in z produces a slight leftwards shift of the $\dot{E} = 0$ locus, and strongly raises its vertical asymptote. In many cases this would produce a fall in E_j . However, when we transform S_j into a parameter that is invariant with respect to prices, we get a well behaved steady state quantity E_j as a function of prices and parameters (including S_j)⁶:

$$G(E_j, S_j, z) = ((\rho + g_2 S_j) S_j - E_j) \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right) - \frac{\alpha}{z} S_j^{-\alpha} = 0 \quad (5)$$

In Appendix 2, as part of proving that this fossil fuel sector satisfies the requirements of the abstract energy sector set out in section 2, it is shown that this relationship implies that the output, E_j , of a firm operating at energy quality, S_j , is an increasing function of the price, z .

The marginal firm in this sector will be the firm that places zero value of operating or not

⁶Just to repeat though, the firm operating at S_j in the economy characterised by s_2 is not the same firm that operated at S_j in the economy characterised by s_1 .

operating. Since we are only discussing the steady states of the production of these firms then the zero value condition is the same as the zero profit flow condition i.e. for any given price level, z , there exists a threshold energy quality level:

$$\begin{aligned}\pi_{min} &= p_E \left(E_{min} - \frac{1}{z} \left(E_{min}^{\frac{1}{\beta}} + S_{min}^{-\alpha} \right) \right) = 0 \\ \text{i.e. } S_{min}(z) &= \left(zE_{min} - E_{min}^{\frac{1}{\beta}} \right)^{-\frac{1}{\alpha}}\end{aligned}$$

Where E_{min} is related to S_{min} by Equation (5). The aggregate energy sector consists of a continuum of these firms operating over (S_{min}, s) i.e.

$$\begin{aligned}E &= \int_{S_{min}}^s E_j dS_j \\ Q_E &= \int_{S_{min}}^s Q_j dS_j = \int_{S_{min}}^s \left(E_j^{\frac{1}{\beta}} + S_j^{-\alpha} \right) dS_j \\ \Pi &= \int_{S_{min}}^s \pi_j dS_j = p_E \left(E - \frac{1}{z} Q_E \right)\end{aligned}$$

It can be shown that S_{min} is a decreasing function of z and so again we have a price taking energy sector that uses intermediate goods such that its output responds endogenously to the relationship between the output energy price and the input intermediates price. Exploitation of lower quality resources is an increasing function of the energy to intermediates price ratio. If it is optimal to exploit a particular resource, then it is optimal to exploit every resource of higher quality, and so the available high quality resources are always exploited. Appendix 2 shows that this fossil fuel energy sector satisfies the properties of the generic energy sector specified in section 2 and hence that the aggregate economy has the steady state behaviour of section 3.

Finally, we consider the behaviour of an infinitesimal (so does not affect the rest of the economy), price taking fossil fuel firm managing a truly non renewable resource. What are the incentives for such a firm as the aggregate economy operates at progressively lower levels of energy quality (though always in the neighbourhood of the steady state if it exists⁷)? With constant or weakly increasing returns to scale in the intermediate goods sector, the firm will expect increasing profits (per unit extracted from a given quality of resource) and so will tend to defer extracting resources. Any resources that are not profitable to extract now, will become profitable to extract as the energy

⁷i.e. we consider some exogenous dynamic process for s_t such that $\dot{s}_t < 0$, $z_t \approx z^*(s_t)$ and $r_t \approx \rho$. This putative infinitesimal non renewable resource firm then manages its finite stock of resources under a variable price regime, but the problem is now tractable because it is partial equilibrium since its decisions do not feed back into the economy-wide prices.

quality exploited in the rest of the economy declines. This is in keeping with the sanguine view of non renewable resource economics since Hotelling. However, with strongly increasing returns to scale in the intermediate goods sector, the firm will expect decreasing profits (per unit extracted from a given quality of resource) and so will tend to bring forward the extraction of resources. Any resources that are not profitable to extract now, will never become profitable to extract as the energy quality exploited in the rest of the economy declines. This is a new result not at all in keeping with the usual picture from the non renewable resource literature in economics. Eventually if the economy can no longer maintain a steady state, the interest rate will fall below the rate of time preference. In this circumstance there is some incentive for profitable firms to defer extracting resources since, even though profits (per unit extracted from a given quality of resource) are decreasing, the value placed on future profits is rising as the interest rate falls. It remains the case however that currently unprofitable resources will never be brought 'on-stream'.

6 Illustrative Results

In this section I present illustrative results generated from the model described in sections 2, 3 & 4 i.e. the full macroeconomy with a renewable energy sector. These results illustrate the mechanism described, but strong increasing returns to scale is contingent upon an extreme parameterisation. To see if it is possible that the phenomena described in this article have any possibility of being quantitatively important, I extend the intermediate goods sector to a monopolistically competitive industry with symmetric firms making heterogenous intermediate goods by renting the capital stock and consuming energy services (i.e. two inputs as opposed to the single input of capital stock in the basic model). These heterogenous goods are aggregated for use in the energy sector and the final goods sector using Dixit-Stiglitz aggregation. This generalised model embeds the basic model⁸ and exhibits the same phenomena (but clearly the propositions derived in section 3 only strictly apply to the basic model).

The results are presented here as plots of the steady state value of the index, r , of the marginal, zero profit, energy firm against the energy quality parameter (which is the index of the energy firm exploiting the highest quality resources), $e = 1/s$. Therefore an increase in e on the x-axis corresponds to a reduction in energy quality, and so this is a plot of a comparative static (at steady state) across a continuum of different economies, each having a different energy quality parameter. Different lines on the plot show this comparative static for economies with different levels, ψ , of scale economies in their intermediate goods sectors. The normalisation constant, θ ,

⁸The basic model is isomorphic to the generalised model with an elasticity of intermediate sector output with respect to energy equal to zero. However, the generalised model has been set up in a more specific, less general, way, with the microfoundations of monopolistic competition under CES aggregation, rather than just assuming zero profits.

in the intermediate goods sector production function is chosen as a function of the degree of scale economies, $\theta = \theta(\psi)$, so that for $e = 0$ (i.e. $s \rightarrow \infty$) the steady state values for the endogenous variables in the model are independent of the level of scale economies, ψ (and therefore the lines on the plot start from a common value, r_0 at $e = 0$) so that the results can be compared on a single plot. If the index of the marginal firm rises (i.e. lower quality resources are exploited) as e rises (i.e. as energy quality falls) then relative energy prices are rising with falling energy quality and the exploitation of lower quality resources expands to (partially) offset the loss of the high quality resources. However, if r falls as e rises then relative energy prices are falling with falling energy quality, and the economy is heading for collapse (in the sense discussed in section 3). This is shown in Figure 6.1 which charts how the steady state of the economy varies with ψ and e . Figure 6.1 does not show timepaths, but we can imagine that if a specific ψ line shows the path of an economy which undergoes a series of depletion shocks⁹, then a rising $r(e)$ curve indicates that the use of lower quality resources substitutes for the high quality resources that are no longer available to the economy. A falling $r(e)$ curve indicates that marginal resources are abandoned as high quality resources cease to be available. The parameters used to generate these results, and the results from the following section, are listed in Appendix 3.

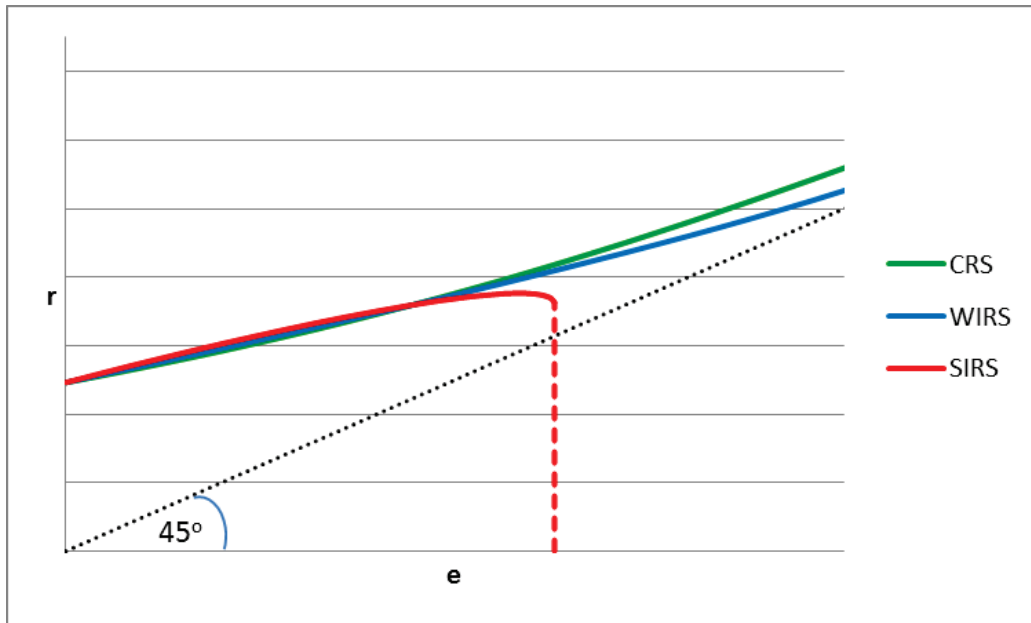


Figure 6.1: $r(e)$ for 3 economies: CRS with $\psi = 1$, WIRS with $1 < \psi < \psi^*$, and SIRS with $\psi > \psi^*$.

Under SIRS, we observe that, initially, the (infinitely) abundant low quality resources provide a substitute for unavailable high quality resources. Eventually as e continues to rise, the $r(e)$ curve

⁹With a point on the graph only generated once the economy has converged to steady state following each depletion shock.

has a turning point and as e rises further the economy starts to abandon the marginal resources despite their abundance, and ultimately the economy collapses. This can be seen more clearly by zooming into figure 6.1 as is shown in Figure 6.2.

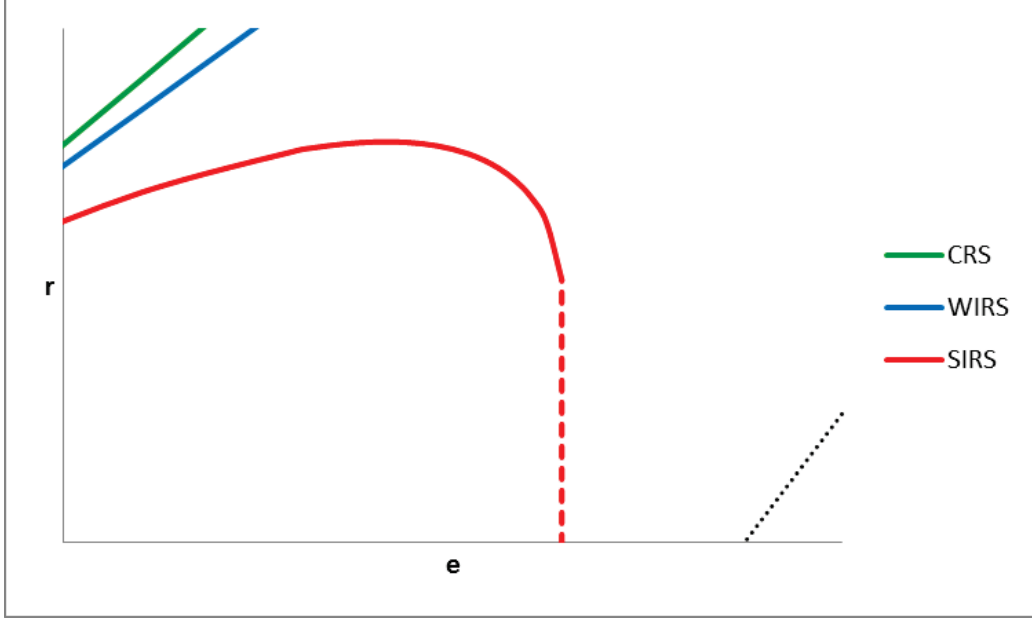


Figure 6.2: As Figure 6.1 but zoomed in.

These results are somewhat unsatisfactory since proposition 4 gives us that $\psi^* = 1/(\gamma + \mu)$. This implies that, given standard estimates from incomes shares, $\psi^* \approx 3$ - vastly higher than any plausible estimate of the degree of scale economies in the real world. Determining whether the phenomena described in this article is an irrelevant feature which real world parameters do not remotely approach, or whether it is worthy of investigating quantitatively, is the purpose of the following model generalisation which essentially adds an energy input requirement to the operation of the intermediate goods sector.

The intermediate goods sector is generalised by splitting it into two. A perfectly competitive *aggregation* sector buys the output of a monopolistically competitive sector which produces heterogenous goods. The aggregation sector has production function and profits as follows:

$$Q + Q_E = \left(\int_0^n q_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1$$

$$\pi_A = p_Q (Q + Q_E) - \int_0^n p_i q_i di = 0$$

The monopolistically competitive sector consists of measure n (endogenous) firms each producing a differentiated good with some monopoly pricing power. The demand schedule that each monopolist

faces, and their production and profit functions are:

$$\begin{aligned} q_i &= (Q + Q_E) p_i^{-\sigma} p_Q^\sigma \\ q_i &= \phi(E_i^\eta K_i^{1-\eta} - f) \\ \pi_i &= p_i q_i - p_E E_i - \rho K_i \end{aligned}$$

Where f is a fixed cost. There is free entry so the profits of each monopolist are driven to zero. The equilibrium conditions for this sector are:

$$\begin{aligned} E_Q &= \left(\frac{\eta}{1-\eta} \frac{p_K}{p_E} \right) K \\ p_Q &= \frac{\sigma}{\sigma-1} \frac{p_K}{\phi(1-\eta)} \left(\frac{\eta}{1-\eta} \frac{p_K}{p_E} \right)^{\frac{\eta\sigma}{1-\sigma}} (\sigma f)^{\frac{1}{\sigma-1}} K^{\frac{1}{1-\sigma}} \\ Q + Q_E &= (\sigma-1) \phi f \left(\frac{\eta}{1-\eta} \frac{p_K}{p_E} \right)^{\frac{\eta\sigma}{\sigma-1}} (\sigma f)^{\frac{\sigma}{1-\sigma}} K^{\frac{\sigma}{\sigma-1}} \end{aligned}$$

The only other change from the model presented previously is that the total output of the energy sector now has to be split across the final and intermediate goods sectors, $E + E_Q$. The fixed cost, f here performs normalisation role as the parameter θ in the basic model: defining $f = f(\sigma)$ allows us to normalise the economies with different σ 's so that they all coincide for $e = 0$. Simulating this model (with the parameters detailed in Appendix 3) produces very similar results to the basic model (see Figures 6.3 & 6.4) but now, as detailed in Appendix 3, the degree of returns to scale needed for SIRS and collapse is much lower¹⁰ than in the basic model¹¹. This suggests that the SIRS mechanism is not obviously ruled out by the parameterisation needed to observe it, and so this is a phenomenon that is worthy of quantitative investigation.

7 Policy & Innovation

The collapse of a SIRS economy as energy quality declines is due to prices and not to any fundamental limits. By construction, low quality resources are infinitely abundant, and a high relative energy price will ensure that they are profitable to exploit. An energy subsidy will therefore bring resources into production and will increase the scale of intermediates sector, improving the allocation across the economy. The economy does not collapse under CRS, and the market allocation

¹⁰Depending of the value of the parameter η . The generalised model exhibits SIRS so long as $\sigma < \sigma^* = 1/((1-\eta)(1-\mu-\gamma))$ - see Appendix 4.

¹¹The SIRS result, as described in Appendix 3, is generated with a returns to scale, $\psi = 1.3 \iff \sigma = 4.3333$, which is the number estimated by Caballero and Lyons (1989).

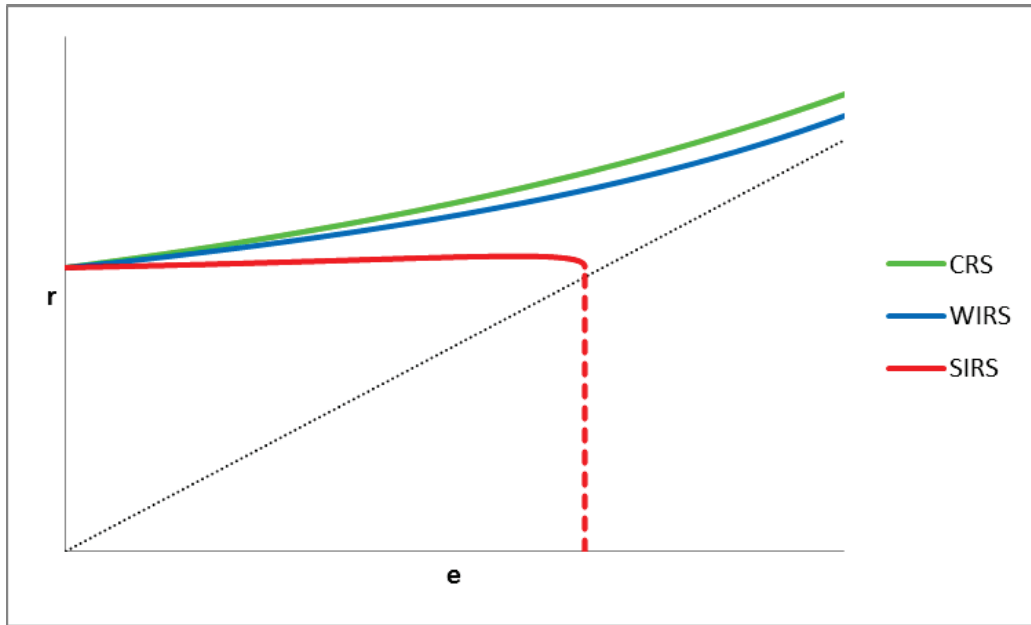


Figure 6.3: $r(e)$ for 3 economies: CRS, WIRS, & SIRS; from the generalised model.

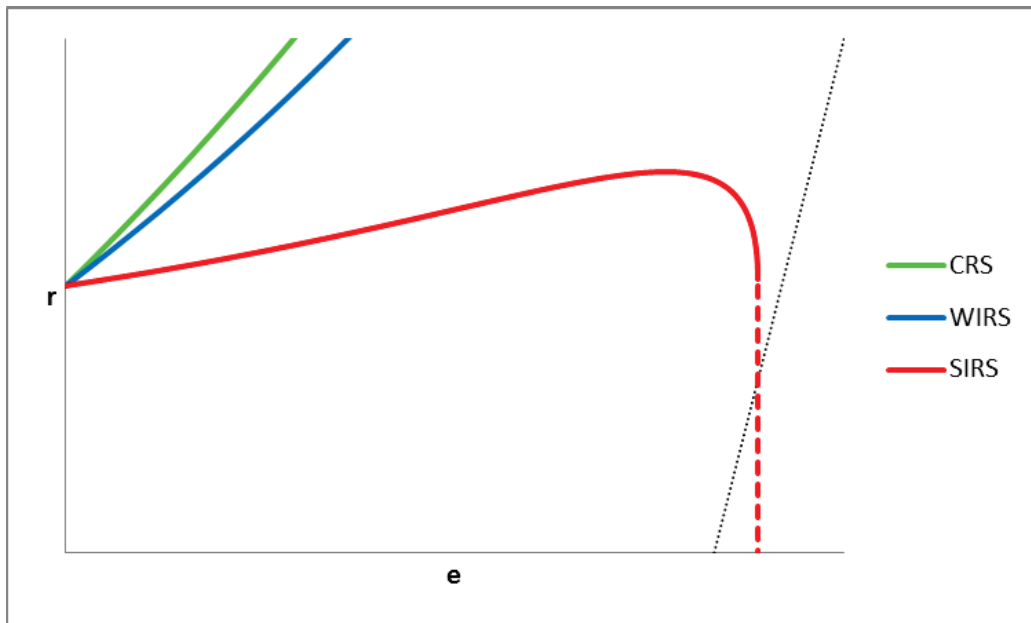


Figure 6.4: As Figure 6.3 but zoomed in.

cannot be improved upon¹². If we were to impose lump sum taxes on the households and use the proceeds to subsidise energy production, then the pareto-optimal tax rate is zero for CRS and is increasing in the degree of scale economies: the more at risk of collapse the economy is from the interaction of increasing returns and energy quality, the more amenable this situation is to policy

¹²This follows from The First Welfare Theorem.

intervention.

When we imagine technological solutions to energy scarcity, we often think of high technology goods that use energy more efficiently. This is the situation that [Aghion, Howitt, Brant-Collett, and Garcaia-Peanalosa \(1998\)](#) abstract from in their model of endogenous growth with non-renewable resources. A real world example of this could be modern cars with computer optimised engines, as opposed to simpler vehicles that use petrol inputs much less efficiently. In Aghion & Howitt's model, as energy became scarce and expensive there was an increasing incentive to develop this technology. However, in the model presented in this paper I suggest that, if we live in a world of SIRS, then energy scarcity may not motivate us to use this advanced energy efficient technology, because the price of the high-tech computer optimised capital goods could rise by more than the price of energy, and consumers will substitute away from such goods towards goods that use intermediates efficiently but energy inefficiently.

Specifically, we could formulate two alternative production technologies for final goods: an intermediates intensive technology that uses energy very efficiently, and an energy intensive technology that used intermediates efficiently. There would be some price ratio at which these technologies used energy and capital services in the same ratio to produce the same output level, and this price ratio would be the price ratio that the economy switched from one technology to the other. In the results presented in Section 6, rising e always causes rising r in the WIRS economies, and causes rising r in the SIRS economy initially. Rising r occurs because of a rising energy to intermediates price. The switch price will eventually be reached for the WIRS economies and they will ultimately use the energy efficient technology. The switch price may or may not be reached under SIRS, but even supposing that it is, further declines in high quality energy availability could see the switch price being reached again on the way down i.e. the SIRS economy may choose never to take up the energy efficient technology, and even if the economy does adopt it, it may then abandon it. This is intuitive - we can well imagine that productivity in advanced sectors depends on sufficient scale, and if scale is hit hard enough by a shortage of energy, then advanced energy efficient products may not be available.

The same issue arises for a putative backstop technology. We could suppose that some non-depletable backstop was available at some relative energy price. For expositional purposes let us suppose it is large scale deployment of solar panels in deserts. Again the productivity of the sectors that can produce the solar technology depends upon the scale at which it operates. At low levels of scarcity there is large demand for semiconductor technology so this sector is large and productive. It appears that the solar backstop is feasible but the relative energy price is too low to justify its deployment. Energy scarcity rises and energy prices rise (relative to wages). However intermediate goods sectors across the economy contract and productivity falls. The pricing is such that despite

the rise in energy prices, we are no closer to profitably deploying the backstop technology. The economy eventually collapses for lack of energy, and at no point was it profitable to deploy the backstop technology. The description here only applies to the SIRS economy, under WIRS, the backstop technology will eventually be deployed.

In general, this story applies to any innovation effort that may allow an economy to grow or continue at the same level under resource restrictions. If the benefits to innovating are positively related to the energy price, but the costs are positively related to the intermediates price, then it will eventually be optimal to undertake the innovation effort under WIRS. It may be the case that it is never be optimal to undertake the innovation effort under SIRS. This problem is amenable to policy intervention though: subsidies can support the scale of industry so that innovations or technologies are within the reach of a SIRS economy, whereas they may be out of reach without policy interventions. This is therefore (theoretical) support (though not necessarily support in any specific case, or for our *real world* economy) for subsidies, e.g. renewables feed-in-tariffs, which may create an industry of sufficient scale to be profitable.

8 Conclusion

I find that the price movements caused by declining energy resources may not be conducive to the exploitation of more marginal resources. Such price movements can lead to macroeconomic collapse, for lack of energy, before all the technologically available energy resources have been exploited. This result is in contrast with the basic Hotellings model and almost all of the non-renewable natural resources and energy literature. Increasing returns to scale, as estimated as occurring in, and often assumed for, manufacturing sectors and industrial economies, is a sufficient condition for this phenomena to be manifest. Innovative or technological solutions to future energy shortages are also adversely affected by this phenomena. However, the more that this phenomena is a real problem, the more it is amenable to policy intervention - which does allow society to mitigate the problem through activist policy.

The mechanism underlying this interaction effect between scale economies and energy quality is that energy supply decisions are positively related to the energy price, and negatively related to input prices. Scale economies can cause productivity in the intermediate sectors, that manufacture inputs for the energy sector, to fall as their scale falls. This can mean that an energy quality shock, which is a supply shock to the whole economy, causes the supply of intermediates to fall by more than the energy supply, and so the scarce commodity is the intermediates. Hence the energy price rise is less than the intermediates price rise and energy sector profitability falls. The existing literature only considers the positive relationship of energy supply with the energy price, and does

not consider intermediate inputs at all.

This mechanism may be quantitatively important: the recent experience of historically high energy prices (likely caused by the supply of high quality, low marginal cost, oil resources failing to rise to match rising demand, predominantly from China, see e.g. [Kilian and Murphy \(2010\)](#)) has not led to a uniform pursuit of alternative energy resources. Renewables investment has been volatile¹³, worldwide nuclear electricity generation has seen absolute declines¹⁴, and for OECD countries the recent experience of high energy prices is associated with demand destruction rather than supply increases¹⁵. Obviously there are multiple reasons for these outcomes such as the short run demand side effects of the global financial crisis, as well as policy decisions due to the Fukushima disaster. However, along with high energy prices there have also been some price increases in the industries which supply the renewable energy industry due to higher commodity costs and supply bottlenecks¹⁶. Energy sector costs rising with the energy price, such that the profitability of marginal suppliers not necessarily improving with a scarcity induced rise in the energy price, is the basic mechanism underlying this article; and so real world experience may be consistent with the phenomena described by this paper.

Future research must develop techniques for testing whether the interaction of scale economies and energy quality is quantitatively important. What are the returns to scale of real manufacturing sectors, especially of those sectors which supply components for extractive industries? Do, for example, growth accounting exercises suggest that we live in a world of constant or weakly increasing returns to scale, or do we live in a world of strongly increasing returns to scale?

¹³See e.g. [The GWEC \(2011\)](#).

¹⁴See e.g. [EPI \(2012\)](#).

¹⁵From [EIA \(2012\)](#) OECD petroleum consumption fell by 10% from 2007 to 2011.

¹⁶See figure 0.3 of [EWEA \(2009\)](#) which shows cost reductions from 1987 to 2004 as technology improved, with cost increases for the final data point in 2006.

References

- AGHION, P., P. HOWITT, M. BRANT-COLLETT, AND C. GARCAIA-PEANALOSA (1998): *Endogenous Growth Theory*.
- BARNETT, H., AND C. MORSE (1963): *Scarcity and Growth*. Johns Hopkins Univ Pr.
- BASU, S., AND J. FERNALD (1997): “Returns to Scale in U.S. Production: Estimates and Implications,” *Journal of Political Economy*.
- BELTRATTI, A., G. CHICHILNISKY, AND G. HEAL (1998): “Sustainability: Dynamics and Uncertainty,” chap. Sustainable Use of Renewable Resources. Springer.
- BRANDT, A., AND A. FARRELL (2007): “Scraping the bottom of the barrel: greenhouse gas emission consequences of a transition to low-quality and synthetic petroleum resources,” *Climatic Change*.
- CABALLERO, R., AND R. LYONS (1989): “The Role of External Economies in U.S. Manufacturing,” *NBER Working Paper*.
- DASGUPTA, P., AND G. HEAL (1974): “The optimal depletion of exhaustible resources,” *Review of Economic Studies*.
- EIA (2012): “O.E.C.D. Petroleum Consumption Data,” Data query, The US Energy Information Administration, www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=5&pid=5&aid=2&cid=CG5,&syid=2007&eyid=2011&unit=TBPD.
- EPI (2012): “World Nuclear Electricity Generation Down 5 Percent Since 2006,” Data highlights, The Earth Policy Institute, www.earth-policy.org/data_highlights/2012/highlights33.
- EWEA (2009): “The Economics of Wind Energy,” Report, European Wind Energy Association, www.windenergie.nl/sites/windenergie.nl/files/documents/the_economics_of_windenergy_ewea.pdf.
- GWEC (2011): “Global Wind Report 2010,” Report, The Global Wind Energy Council, www.gwec.net/publications/global-wind-report-2/global-wind-report-2011/.
- HALL, R. (1989): “Invariance Properties of Solow’s Productivity Residual,” *NBER Working Paper*.
- HAMILTON, J. (2011): “Oil Prices, Exhaustible Resources, and Economic Growth,” *Prepared for ‘Handbook of Energy and Climate Change’*.

- HOLLAND, S. (2008): “Modeling Peak Oil,” *The Energy Journal*.
- HOTELLING, H. (1931): “The Economics of Exhaustible Resources,” *The Journal of Political Economy*.
- KILIAN, L., AND D. MURPHY (2010): “The role of inventories and speculative trading in the global market for crude oil,” *CEPR Discussion Papers*.
- MOHLER, L., AND M. SEITZ (2010): “The Gains from Variety in the European Union,” *Munich Discussion Paper* <http://epub.ub.uni-muenchen.de/11477/>.
- MURPHY, D., AND C. HALL (2010): “Year in review - EROI or energy return on (energy) invested,” *Annals of The New York Academy of Sciences*.
- MURPHY, K., A. SCHLEIFER, AND R. VISHNY (1989): “Industrialization And The Big Push,” *Journal of Political Economy*.
- MURPHY, T. (2011): “Peak Oil Perspective,” Blog post, physics.ucsd.edu/do-the-math/2011/11/peak-oil-perspective.
- SIMON, J. L. (1996): *The Ultimate Resource 2*. Princeton University Press.
- VENTURA, J. (2005): “A global view of economic growth,” chap. in: P. Aghion and S. Durlauf (ed.) “Handbook of Economic Growth”.

A Appendix 1: The renewable energy sector satisfies requirements of section 2

Throughout we refer to:

$$\begin{aligned}
 e(s) &= s^{-1} \\
 r(z) &= (\beta z)^{\frac{\beta}{1-\beta}} (1-\beta) \\
 E &= \frac{1}{2} (1-\beta^2) (\beta z)^{\frac{2\beta}{1-\beta}} - e(\beta z)^{\frac{\beta}{1-\beta}} + \frac{1}{2} e^2 \\
 Q_E &= (1-\beta) (\beta z)^{\frac{1+\beta}{1-\beta}} - e(\beta z)^{\frac{1}{1-\beta}}
 \end{aligned}$$

Proposition A. 1. 1. $\exists z_{min}(e)$ s.t. $E(z_{min}(e), e) = 0$ and $Q_E(z_{min}(e), e) = 0$ with $\frac{dz_{min}(s)}{ds} = \frac{dz_{min}(e)}{de} \frac{de}{ds} < 0 \iff \frac{dz_{min}(e)}{de} > 0$ (since $e = s^{-1}$) and $z_{min} \rightarrow \infty$ as $e \rightarrow \infty$.

Proof. z_{min} is defined by the marginal firm being the only firm i.e. $r(z_{min}) = (1-\beta)(\beta z)^{\frac{\beta}{1-\beta}} = e$, so that no energy industry exists.

$$z_{min} = \frac{1}{\beta} \left(\frac{1}{1-\beta} \right)^{\frac{1-\beta}{\beta}} e^{\frac{1-\beta}{\beta}}$$

Plugging this expression into the equations for E and Q_E gives zero as required (though this is by definition), and we can immediately see that z_{min} is an increasing function of e and that $z_{min} \rightarrow \infty$ as $e \rightarrow \infty$. \square

Proposition A. 1. 2. $\frac{\partial E}{\partial z} > 0$

Proof.

$$\begin{aligned}
 \frac{\partial E}{\partial z} &= \frac{\beta^2}{1-\beta} (\beta z)^{\frac{2\beta-1}{1-\beta}} \left((1-\beta^2)(\beta z)^{\frac{\beta}{1-\beta}} - e \right) > 0 \\
 \text{so long as } z &> \frac{1}{\beta} \left(\frac{e}{(1+\beta)(1-\beta)} \right)^{\frac{1-\beta}{\beta}} \\
 \text{true, since: } z &> z_{min} = \frac{1}{\beta} \left(\frac{e}{1-\beta} \right)^{\frac{1-\beta}{\beta}} > \frac{1}{\beta} \left(\frac{e}{(1+\beta)(1-\beta)} \right)^{\frac{1-\beta}{\beta}}
 \end{aligned}$$

\square

Proposition A. 1. 3. $\frac{\partial Q_E}{\partial z} > 0$

Proof.

$$\begin{aligned}\frac{\partial Q_E}{\partial z} &= \frac{\beta}{1-\beta} (\beta z)^{\frac{\beta}{1-\beta}} \left((1-\beta^2)(\beta z)^{\frac{\beta}{1-\beta}} - e \right) > 0 \\ \text{so long as } z &> \frac{1}{\beta} \left(\frac{e}{(1+\beta)(1-\beta)} \right)^{\frac{1-\beta}{\beta}} \\ \text{true, since: } z &> z_{\min} = \frac{1}{\beta} \left(\frac{e}{1-\beta} \right)^{\frac{1-\beta}{\beta}} > \frac{1}{\beta} \left(\frac{e}{(1+\beta)(1-\beta)} \right)^{\frac{1-\beta}{\beta}}\end{aligned}$$

□

Proposition A. 1. 4. $\lim_{z \rightarrow \infty} E = a(e)z^x$, $x \geq 0$

Proof. Clearly $\lim_{z \rightarrow \infty} E = \frac{1}{2}(1-\beta^2)\beta^{\frac{2\beta}{1-\beta}}z^{\frac{2\beta}{1-\beta}}$, so $x = \frac{2\beta}{1-\beta} > 0$ and $a(e) = \frac{1}{2}(1-\beta^2)\beta^{\frac{2\beta}{1-\beta}}$. □

Proposition A. 1. 5. $\lim_{z \rightarrow \infty} \left(\frac{Q_E}{E} \right) = b(e)z^y$, $0 \leq y \leq 1$

Proof.

$$\begin{aligned}\text{Have } \lim_{z \rightarrow \infty} Q_E &= (1-\beta)(\beta z)^{\frac{1+\beta}{1-\beta}} \\ \text{so } \lim_{z \rightarrow \infty} \left[\frac{Q_E}{E} \right] &= \frac{2\beta}{(1+\beta)}z\end{aligned}$$

So $b(e) = \frac{2\beta}{1+\beta}$ and $y = 1 \in [0, 1]$. □

Proposition A. 1. 6. $\lim_{z \rightarrow z_{\min}} E = c(e)g(z)$, $g(z_{\min}) = 0$

Proof.

$$\begin{aligned}\text{Have } z_{\min} &= \frac{1}{\beta} \left(\frac{1}{1-\beta} \right)^{\frac{1-\beta}{\beta}} e^{\frac{1-\beta}{\beta}} \Rightarrow e = (1-\beta)(\beta z_{\min})^{\frac{\beta}{1-\beta}} \\ \text{so } E &= \frac{1}{2}(1-\beta^2) \left((\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{\min})^{\frac{\beta}{1-\beta}} \right) \left((\beta z)^{\frac{\beta}{1-\beta}} - \frac{1-\beta}{1+\beta} (\beta z_{\min})^{\frac{\beta}{1-\beta}} \right) \\ \text{i.e. } \lim_{z \rightarrow z_{\min}} E &= \beta e \left((\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{\min})^{\frac{\beta}{1-\beta}} \right)\end{aligned}$$

So $g(z) = (\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{\min})^{\frac{\beta}{1-\beta}}$, with $g(z_{\min}) = 0$, and $c(e) = \beta e$. □

Proposition A. 1. 7. $\lim_{z \rightarrow z_{\min}} Q_E = d(e)g(z)$, $g(z_{\min}) = 0$

Proof.

$$\begin{aligned}Q_E &= (1-\beta)(\beta z)^{\frac{1}{1-\beta}} \left((\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{\min})^{\frac{\beta}{1-\beta}} \right) \\ \text{i.e. } \lim_{z \rightarrow z_{\min}} Q_E &= \left(\frac{e}{1-\beta} \right)^{\frac{1}{\beta}} \left((\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{\min})^{\frac{\beta}{1-\beta}} \right)\end{aligned}$$

So $g(z) = (\beta z)^{\frac{\beta}{1-\beta}} - (\beta z_{min})^{\frac{\beta}{1-\beta}}$, with $g(z_{min}) = 0$, and $d(e) = \left(\frac{e}{1-\beta}\right)^{\frac{1}{\beta}}$. □

B Appendix 2: The fossil fuel energy sector satisfies requirements of section 2

Lemma A. 2. 1. $\frac{dS_{min}}{dz} < 0$.

Proof. For a given price level, $p_E(1)$ & z_1 , $S_{min}(z_1)$ satisfies:

$$\pi_{min} = p_E(1) \left(E_{min} - \frac{1}{z_1} E_{min}^{\frac{1}{\beta}} - \frac{1}{z_1} S_{min}^{-\alpha} \right) = 0$$

If prices change to $z_2 > z_1$ then clearly the firm operating at S_{min} clearly has the option of keeping quantities of inputs constant, continuing to produce E_{min} , and so make strictly positive profits. Consequently there will be at least one firm with $S_M < S_{min}$ which now finds it profitable to start producing i.e. with a rise in z there is a new, lower quality firm. Equivalently, $\frac{dS_{min}}{dz} < 0$. \square

Proposition A. 2. 1. $\exists z_{min}(s)$ s.t. $E(z_{min}(s), s) = 0$ and $Q_E(z_{min}(s), s) = 0$ with $\frac{dz_{min}(s)}{ds} < 0$ and $z_{min} \rightarrow \infty$ as $s \rightarrow 0$.

Proof. z_{min} is defined by the marginal firm being the only firm i.e. $S_{min}(z_{min}(s)) = s$, so that no energy industry exists. Aggregate industry inputs and outputs are defined as an integral over the range $[S_{min}, s]$. At $z = z_{min}$ this interval has zero length and so $E(z_{min}(s), s) = 0$ and $Q_E(z_{min}(s), s) = 0$ are trivially satisfied.

$$\begin{aligned} \text{Have } S_{min}(z_{min}(s)) &= s \\ \text{so } \frac{dS_{min}(z_{min}(s))}{ds} &= \frac{dS_{min}(z)}{dz} \Big|_{z=z_{min}} \frac{dz_{min}}{ds} = 1 \\ \text{i.e. } \frac{dz_{min}}{ds} &= 1 / \frac{dS_{min}(z)}{dz} \Big|_{z=z_{min}} \end{aligned}$$

Lemma 1 shows that $\frac{dS_{min}}{dz} < 0$ in general, and so in particular, $\frac{dS_{min}}{dz} \Big|_{z=z_{min}} < 0$. Therefore $\frac{dz_{min}(s)}{ds} < 0$ as req. Taking the limit of the Equation (5) for the marginal firm as $s \rightarrow 0$ (the z to use for the marginal firm is z_{min} and the energy output is zero in order to be on the $\dot{S} = 0$ locus at $S_j = s = 0$) gives:

$$\frac{\alpha}{z_{min}} = s^\alpha ((\rho + g_2 s) s) \times 1 \rightarrow 0 \text{ as } s \rightarrow 0 \Rightarrow z_{min} \rightarrow \infty \text{ as } s \rightarrow 0$$

i.e. $z_{min} \rightarrow \infty$ as $s \rightarrow 0$. \square

Lemma A. 2. 2. $G_z \equiv \frac{\partial G}{\partial z} > 0$

Proof.

$$\begin{aligned} G(E_j, S_j, z) &= ((\rho + g_2 S_j) S_j - E_j) \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right) - \frac{\alpha}{z} S_j^{-\alpha} \\ \frac{\partial G}{\partial z} &= \beta ((\rho + g_2 S_j) S_j - E_j) E_j^{\frac{1-\beta}{\beta}} (\beta z)^{-2} + \frac{\alpha}{z^2} S_j^{-\alpha} > 0 \end{aligned}$$

□

Assumption A. 2. 1. *Let*

$$0 < \beta < \frac{1}{1 + \frac{z^*(s_i)}{\alpha} S_j^\alpha E_j \left(1 - (\beta z^*(s_i))^{-1} E_j^{\frac{1-\beta}{\beta}} \right)} = B(i, j)$$

$\forall S_j \in [S_{\min}(z^*(s_i)), s_i]$ and $\forall s_i \in (0, s]$ where E_j is given by $G(E_j, S_j, z^*(s_i)) = 0$ and $z^*(s_i) \equiv$ equilibrium price ratio given the energy quality parameter s_i .

This is an uncontroversial regularity assumption. β is the curvature of the energy production function which is already assumed to be in the interval $(0, 1)$. This assumption just narrows the interval somewhat, since we can easily show that the denominator is greater than 1:

$$G(E_j, S_j, z) = 0 \Rightarrow ((\rho + g_2 S_j) S_j - E_j) \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right) = \frac{\alpha}{z} S_j^{-\alpha} > 0$$

$$\text{Have } \rho - g_{j1} + 2g_2 S_j(t) > 0 \text{ and } g_{j1} = \frac{E_j}{S_j} + g_2 S_j$$

$$\text{Combining } \Rightarrow (\rho + g_2 S_j) S_j - E_j > 0$$

$$\Rightarrow 1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} > 0$$

$$\Rightarrow 0 < B(i, j) = \frac{1}{1 + \frac{z^*(s_i)}{\alpha} S_j^\alpha E_j \left(1 - (\beta z^*(s_i))^{-1} E_j^{\frac{1-\beta}{\beta}} \right)} < 1$$

Lemma A. 2. 3. $G_E \equiv \frac{\partial G}{\partial E_j} < 0$

Proof.

$$G(E_j, S_j, z) = ((\rho + g_2 S_j) S_j - E_j) \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right) - \frac{\alpha}{z} S_j^{-\alpha}$$

$$\frac{\partial G}{\partial E_j} = -1 - \frac{1}{\beta^2 z} E_j^{\frac{1-\beta}{\beta}} \left((\rho + g_2 S_j)(1 - \beta) \frac{S_j}{E_j} - 1 \right)$$

$$\text{i.e. } (\rho + g_2 S_j)(1 - \beta) \frac{S_j}{E_j} > 1 \Rightarrow \frac{\partial G}{\partial E_j} < 0$$

$$\begin{aligned}
(\rho + g_2 S_j)(1 - \beta) \frac{S_j}{E_j} &> 1 \\
\Rightarrow (\rho + g_2 S_j) S_j (1 - \beta) - E_j &> 0
\end{aligned}$$

$$\begin{aligned}
\text{Have } (\rho + g_2 S_j) S_j - E_j &= \frac{\alpha}{z} S_j^{-\alpha} \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right)^{-1} \\
\text{so } (\rho + g_2 S_j) S_j (1 - \beta) - E_j &= \frac{\alpha}{z} S_j^{-\alpha} \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right)^{-1} - \beta (\rho + g_2 S_j) S_j
\end{aligned}$$

$$\begin{aligned}
\text{i.e. } \beta &< \frac{\alpha}{z} S_j^{-\alpha} \left(\left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right) ((\rho + g_2 S_j) S_j) \right)^{-1} \\
&= \frac{1}{1 + \frac{z}{\alpha} S_j^\alpha E_j \left(1 - (\beta z)^{-1} E_j^{\frac{1-\beta}{\beta}} \right)}
\end{aligned}$$

$\Rightarrow (\rho + g_2 S_j) S_j (1 - \beta) - E_j > 0$ i.e. Assumption 1 ensures that $\frac{\partial G}{\partial E_j} < 0$. □

Proposition A. 2. 2. $\frac{\partial E}{\partial z} > 0$

Proof.

$$\begin{aligned}
G(E_j, S_j, z) = 0 &\Rightarrow \frac{\partial E_j}{\partial z} = -\frac{G_z}{G_E} \\
\text{i.e. } \frac{\partial E_j}{\partial z} &> 0 \quad \text{by Lemmas 2 \& 3}
\end{aligned}$$

$$\text{Aggregate energy production, } E(z) = \int_{S_{\min}(z)}^s E_j(z) dS_j$$

Every element of this sum is rising as z rises. The upper limit of the sum is constant with respect to z . The lower limit falls with rising z (by Lemma 1). Therefore we clearly have $\frac{\partial E}{\partial z} > 0$. □

Proposition A. 2. 3. $\frac{\partial Q_E}{\partial z} > 0$

Proof.

$$\begin{aligned}
Q_j(z) &= E_j^{\frac{1}{\beta}} + S_j^{-\alpha} \\
\text{i.e. } \frac{\partial Q_j}{\partial z} &= \frac{1}{\beta} E_j^{\frac{1-\beta}{\beta}} \frac{\partial E_j}{\partial z} > 0 \\
Q_E &= \int_{S_{\min}(z)}^s Q_j(z) dS_j
\end{aligned}$$

Again, every element of this sum is rising as z rises. The upper limit of the sum is constant with respect to z . The lower limit falls with rising z (by Lemma 1). Therefore we clearly have $\frac{\partial Q_E}{\partial z} > 0$. \square

Proposition A. 2. 4. $\lim_{z \rightarrow \infty} E = a(s)z^x, x \geq 0$

Proof. Clearly $\lim_{z \rightarrow \infty} E_j = (\rho + g_2 S_j) S_j$ (from Equation (5)) and $\lim_{z \rightarrow \infty} S_{min} = 0$. Therefore:

$$\lim_{z \rightarrow \infty} E = \int_0^s (\rho S_j + g_2 S_j^2) dS_j = \frac{1}{2} \rho s^2 + \frac{1}{3} g_2 s^3$$

i.e. $a(s) = \frac{1}{2} \rho s^2 + \frac{1}{3} g_2 s^3$ and $x = 0$. \square

Proposition A. 2. 5. $\lim_{z \rightarrow \infty} \left(\frac{Q_E}{E} \right) = b(s)z^y, 0 \leq y \leq 1$

Proof.

$$\begin{aligned} \text{Have } Q_j &= E_j^{\frac{1}{\beta}} + S_j^{-\alpha} \rightarrow (\rho S_j + g_2 S_j^2)^{\frac{1}{\beta}} + S_j^{-\alpha} \text{ as } z \rightarrow \infty \\ \text{so } \lim_{z \rightarrow \infty} \left[\frac{Q_E}{E} \right] &= \frac{\int_0^s ((\rho S_j + g_2 S_j^2)^{\frac{1}{\beta}} + S_j^{-\alpha}) dS_j}{\frac{1}{2} \rho s^2 + \frac{1}{3} g_2 s^3} = b(s) \end{aligned}$$

So $b(s)$ is as above and $y = 0 \in [0, 1]$. \square

Proposition A. 2. 6. $\lim_{z \rightarrow z_{min}} E = c(s)g(z), g(z_{min}) = 0$

Proof.

$$\lim_{z \rightarrow z_{min}} E = E_s(z_{min}) \times (s - S_{min}(z))$$

So $g(z) = s - S_{min}(z)$ gives the result. $E_s(z_{min})$ is a strictly positive constant giving the energy output defined by the zero profit condition for the firm operating at s . \square

Proposition A. 2. 7. $\lim_{z \rightarrow z_{min}} Q_E = d(s)g(z), g(z_{min}) = 0$

Proof.

$$\lim_{z \rightarrow z_{min}} Q_E = Q_s(z_{min}) \times (s - S_{min}(z))$$

So $g(z) = s - S_{min}(z)$ gives the result. $Q_s(z_{min})$ is a strictly positive constant giving the intermediate inputs defined by the zero profit condition for the firm operating at s . \square

C Appendix 3: Parameters of simulated economies

Parameter	Basic Model	Generalised Model
A	1	1
β	0.5	0.5
μ	0.111111111	0.04
γ	0.222222222	0.293333333
δ	0.1	0.1
ρ	0.05	0.05
η	n/a	0.7
φ	n/a	1
ψ	Used	Equivalent
CRS	1	1
WIRS	2.9	1.111111111
ψ^*	3	1.25
SIRS	3.1875	1.3
ψ^{**}	3.375	1.777251185
σ	Equivalent	Used
CRS	∞	∞
WIRS	1.526315789	10
σ^*	1.5	5
SIRS	1.457142857	4.333333333
σ^{**}	1.421052632	2.286585366

D Appendix 4: SIRS in generalised model

Start with zero profits in the intermediate goods market and substitute:

$$\begin{aligned}
p_K K + p_E E_Q &= p_Q(Q + Q_E) \\
\text{i.e. } \frac{p_K K}{1 - \eta} &= p_Q(Q + Q_E) \\
\text{but } K &= \sigma f [(\sigma - 1)\phi f]^{\frac{1-\sigma}{\sigma}} \left(\frac{\eta}{1 - \eta} \frac{p_K}{p_E} \right)^{-\eta} (Q + Q_E)^{\frac{\sigma-1}{\sigma}} \\
\text{so } p_K &= \frac{1 - \eta}{\sigma f} [(\sigma - 1)\phi f]^{\frac{\sigma-1}{\sigma}} \left(\frac{\eta}{1 - \eta} \frac{p_K}{p_E} \right)^{\eta} (Q + Q_E)^{\frac{1-\sigma}{\sigma}} p_Q(Q + Q_E) \\
&= \frac{1 - \eta}{\sigma f} [(\sigma - 1)\phi f]^{\frac{\sigma-1}{\sigma}} \left(\frac{\eta}{1 - \eta} \right)^{\eta} \left(\frac{p_K}{z} \right)^{\eta} p_Q^{1-\eta} (Q + Q_E)^{1+\frac{1-\sigma}{\sigma}} \\
\text{i.e. } p_K^{1-\eta} &= \frac{1 - \eta}{\sigma f} [(\sigma - 1)\phi f]^{\frac{\sigma-1}{\sigma}} \left(\frac{\eta}{1 - \eta} \right)^{\eta} z^{-\eta} p_Q^{1-\eta} (Q + Q_E)^{\frac{1}{\sigma}} \\
\text{i.e. } p_K &= c_0 z^{\frac{\eta}{\eta-1}} p_Q(Q + Q_E)^{\frac{1}{\sigma(1-\eta)}} \\
&= c_1 z^{\frac{\eta}{\eta-1} + \gamma - 1 + \frac{1}{\sigma(1-\eta)}} E^{\mu + \gamma - 1 + \frac{1}{\sigma(1-\eta)}} \left(\frac{\gamma}{\mu} + \frac{Q_E}{zE} \right)^{\frac{1}{\sigma(1-\eta)}}
\end{aligned}$$

Taking limits:

$$\begin{aligned}
\lim_{z \rightarrow z_{min}} p_K &= c_2 E (z_{min})^{\mu + \gamma - 1 + \frac{1}{\sigma(1-\eta)}} \\
\text{i.e. } \lim_{z \rightarrow z_{min}} p_K = 0 &\text{ if } \mu + \gamma - 1 + \frac{1}{\sigma(1-\eta)} > 0 \\
&\text{i.e. if } \sigma < \frac{1}{(1-\eta)(1-\gamma-\mu)} \\
&\text{i.e. } \sigma^* = \frac{1}{(1-\eta)(1-\gamma-\mu)}
\end{aligned}$$

Likewise:

$$\begin{aligned}
\lim_{z \rightarrow \infty} p_K &= c_3 z^{\frac{\eta}{\eta-1} + \gamma - 1 + \frac{1}{\sigma(1-\eta)} + x(\mu + \gamma - 1 + \frac{1}{\sigma(1-\eta)})} \\
\text{i.e. } \lim_{z \rightarrow \infty} p_K = 0 &\text{ if } \frac{\eta}{\eta-1} + \gamma - 1 + \frac{1}{\sigma(1-\eta)} + x \left(\mu + \gamma - 1 + \frac{1}{\sigma(1-\eta)} \right) < 0 \\
&\text{i.e. if } \sigma > \frac{1}{\frac{\eta}{1+x} + (1-\eta)(1-\gamma - \frac{x}{1+x}\mu)} \\
&\text{i.e. } \sigma^{**} = \frac{1}{\frac{\eta}{1+x} + (1-\eta)(1-\gamma - \frac{x}{1+x}\mu)}
\end{aligned}$$

1 Introduction

In Catalanian local elections held in November 2012, political parties in favour of independence from Spain gained most seats. Catalanian independence is therefore a serious possibility in the short run. Whilst there are many arguments for and against independence that are non-economic in nature, an economic argument is to the fore in the debate in Catalonia: it is generally accepted that this relatively wealthy autonomous community, which pays a large fiscal transfer to the rest of Spain, would be better off on achieving independence as then it would not pay this fiscal transfer. This paper challenges this assumption by undertaking a quantitative analysis of trade costs of the breakup of the Spanish union. The results of this analysis are that these costs are possibly larger than any benefits that Catalonia can expect to realise on independence. At a minimum, in the debate around the future of Catalonia's constitutional position, it should not be taken as given that an economically rational self-interested Catalanian voter would choose independence.

This paper focuses entirely on two aspects of the economic effects of independence or union for Catalonia: trade costs that come with independence; and fiscal transfer costs that come with union. It is relatively easy to measure these redistributive fiscal transfer costs, but it is difficult to measure the advantages of integration. Clearly, there are many other economic issues that we do not consider here e.g. the costs and failures that may arise from overly distant government under union (e.g. [Alesina, Spolaore, and Wacziarg \(2005\)](#)); the costs of “race to the bottom” fiscal competition (with independence); and other sources of economies and dis-economies of scale. Here we only consider the application of new trade theory, following [Anderson and van Wincoop \(2003\)](#), [Melitz \(2003\)](#) and [Arkolakis, Costinot, and Rodríguez-Clare \(2012\)](#), to the quantification of the value of the trade links that come from an integrated economy with a larger market. These trade benefits can arise for many different underlying reasons, for example: a love of variety means that the available product range expands with the size of the market and leads to aggregate increasing returns to scale, as in [Krugman \(1980\)](#); a larger market can lead to better firm selection as efficient firms expand to serve this larger market, putting upward pressure on wages, and lowering profitability of low productivity firms who exit, as in [Melitz \(2003\)](#); & traditional Ricardian trade explanations as in [Eaton and Kortum \(2002\)](#). [Arkolakis, Costinot, and Rodríguez-Clare \(2012\)](#) show that the microfoundations underlying gains from trade, conditional on the value of the elasticity of trade flows to trade frictions, do not affect the calibrated value of these gains. Therefore without loss of generality¹, we can select a specific model to work with. In this paper we develop a version of the [Melitz \(2003\)](#) model that is easy to calibrate.

In order to quantify the benefits of union we need a counterfactual scenario that describes independence. To do this, we suppose that Catalonia will become a country that is as isolated from

¹Though to the extent that the selected model determines the value of the elasticity of trade flows to trade frictions, we do lose generality.

the rest of Spain, in trade terms, as those other independent countries that the data suggests are closest to Spain. We call this trade isolation the *effective distance*. It is difficult to forecast changes in this distance following national separation: we do not have a theory of its determinants. If we did have such a theory, we would use it, but in its absence, we construct a plausible counterfactual scenario. What would happen if the distance between Catalonia and the Rest of Spain came to be the current distance between Spain and Spain's closest trading partner: Portugal? It is not outlandish to think that in the long run Spain's closest trading partner would be a good model for the interaction between Catalonia and the rest of Spain.

We do not try to capture transition dynamics because in the short and medium term it is difficult to guess the degree of interaction, as two forces operate in different directions. The process of separation may be expected to create tensions which would reduce the interactions between the former partners. On the other hand, history must have build strong links that may persist for some time. We do not know how long this transition period will last. Therefore we focus solely on long run steady state in which Catalonia's distance with the rest of Spain is the current distance between Spain and Portugal. The obvious criticism of this approach is that the flow benefits of not paying the fiscal transfer are realised immediately whereas the costs may not be realised for several generations. However, it is possible that the short run costs are higher than the long run costs due to disputes, boycotts and general bad feeling. The long run comparison therefore is akin to a central estimate of the costs because of the uncertain nature of these two opposing forces.

The paper is structured as follows. In section 2 we present empirical evidence that motivates the exercise: we first describe evidence for the extraordinarily high integration of subnational entities (regions, countries, states or nations are all loaded terms!) like Catalonia, relative to independent countries. This evidence motivates our claim that we cannot expect Catalonia to remain as integrated with the Spanish economy post independence, and that increases in trade with the rest of the world will not fully compensate for this change. To the extent that there are gains from trade, this reduction in economic integration may be more or less costly. We then describe the evidence that Portugal is Spain's closest trading partner and is the most suitable counterfactual comparison to make to model an independent Catalonia. In section 3 we develop the model, based on [Melitz \(2003\)](#) and describe the calibration procedure. Section 4 presents the data to which we calibrate. Section 5 describes our results, and section 6 concludes.

2 Empirical Evidence

2.1 Catalonia does not look like a country

Looking at Catalonia's imports and exports naively, it would appear that Catalonia is highly integrated into the global economy with high imports and exports as a percentage of GDP. According to the Centre Català de negocis Catalonia is the third largest exporter as a percentage of GDP in the world.

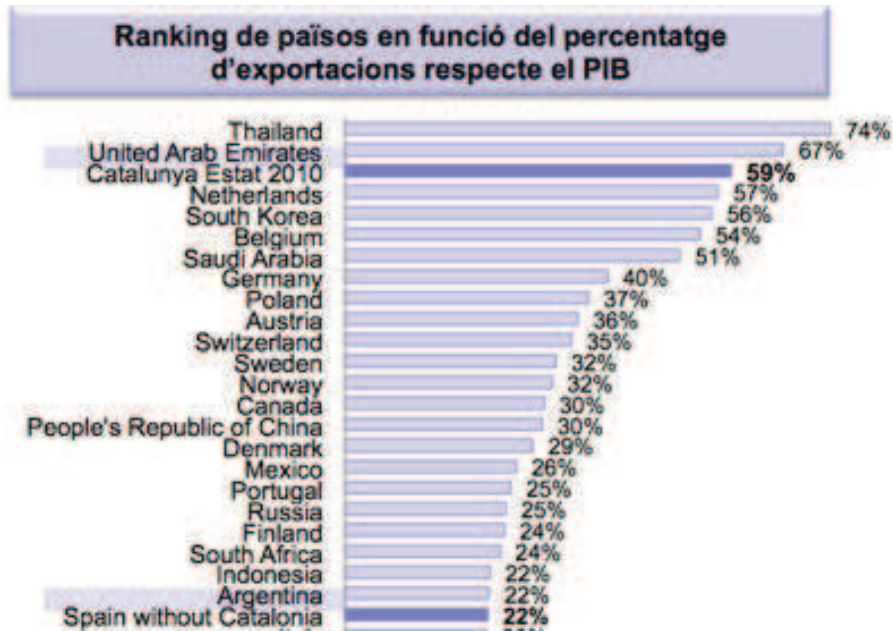


Figure 2.1: The largest exporters in the world as a percentage of GDP. From <http://media.elenoticies.com/ext/20120706/escr.pdf>

However, Catalonia's trade is very concentrated with the rest of Spain: more than 50% of the total external trade is with the rest of Spain rather than internationally. The only country that has a similar trade concentration with a single partner is Canada, because it concentrates its trade with the USA. The USA is vastly larger than Spain though, and if Catalonia really was an open economy then it would be expected to do more trade with France than it does with the rest of Spain since France is larger than the rest of Spain and both are adjacent to Catalonia. In order to quantify these trade concentration facts, we construct a Herfindahl Index of trade concentration².

²If there are N countries, with the exports from country i to country j denoted X_j^i ($X_i^i \equiv 0$), then the Herfindahl Index for country i , $H_i = \sum_{j=1}^N [(X_j^i / \sum_{k=1}^N [X_k^i])^2]$. $H_i = 1$ indicates complete concentration of trade with a single trading partner. $H_i \rightarrow 0$ (equality only possible with infinitely many possible trading partners) indicates complete diversification of trade across all partners.

The following figures highlight how anomalous Catalonia’s trade concentration is compared with independent countries in the European Union. Figure 2.2 shows the Herfindahl index ordered from lowest (diverse trade) to highest (concentrated trade). Catalonia’s Herfindahl Index is approximately double that of the most trade concentrated independent EU member³: it’s an order of magnitude type comparison. Figure 2.3 show the same data but with countries ordered by GDP, since we may expect small countries to trade more, and concentrate this trade with their large neighbours. The negative trend confirms this intuition, and so as a small country we would expect Catalonia to have a high Herfindahl Index: but not nearly as high as we observe.

Catalonia seems to be anomalous as a country. Proponents of independence point to the apparent openness of the Catalonian economy as a feature that will support its prosperity as an independent country. However, the data is suggestive of high Catalonian trade, concentrated with the rest of Spain, being an artifact of the low effective distance that comes from being part of the same country. The conclusion of the calibration exercise undertaken in this paper is that Catalonia concentrates so much its trade with the rest of Spain because its distance with the rest of Spain it is much lower than the normal distance between independent countries. Catalonia’s status as a very open economy is entirely a function of this close integration with the rest of Spain and cannot be expected to survive in the long run after achieving independence.

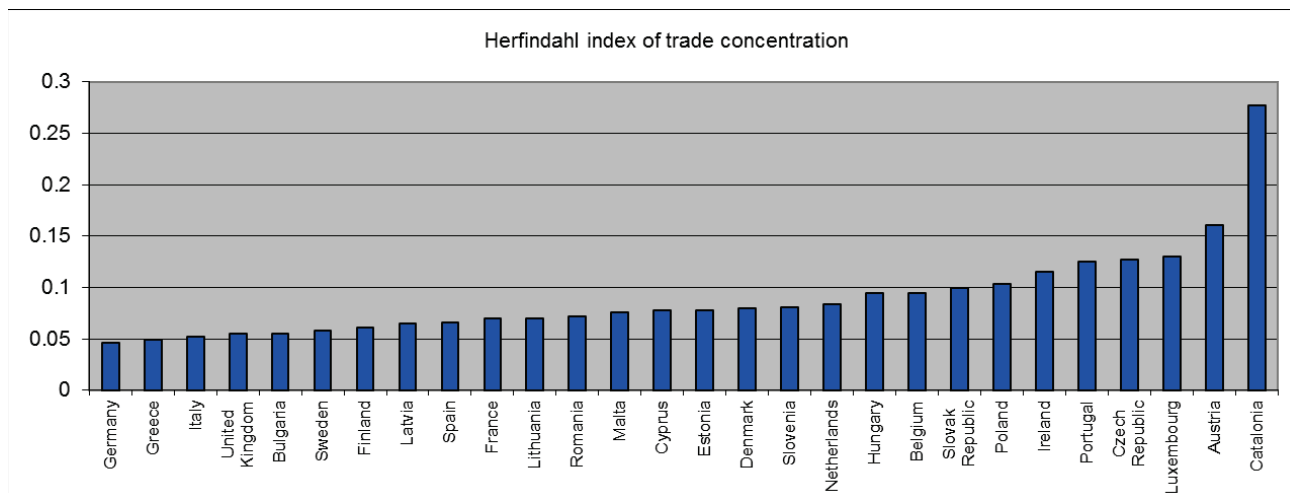


Figure 2.2: Herfindahl Index of EU27, sorted lowest to highest

2.2 The counterfactual should be Portugal

To undertake a cost benefit analysis of Catalonian independence we need to choose a counterfactual to represent the long run distance between Catalonia and the rest of Spain under independence.

³Austria, a small country who concentrate their trade with with the EU’s largest economy, Germany.

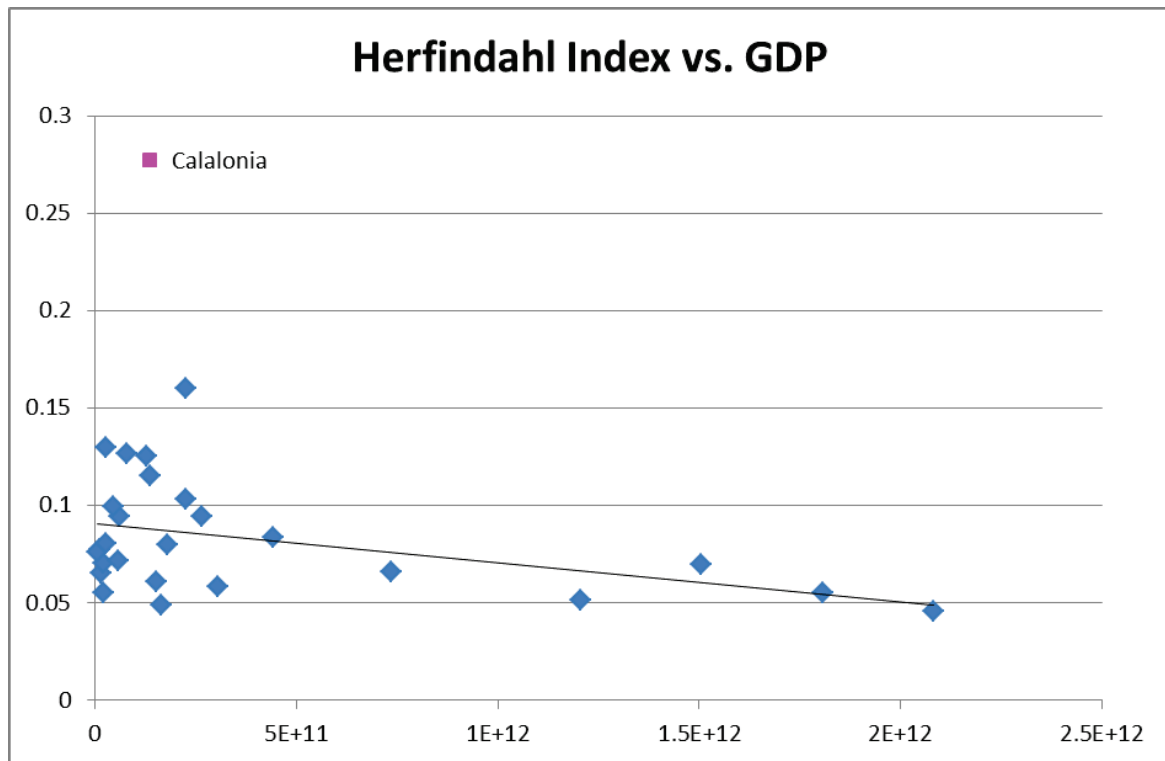


Figure 2.3: Herfindahl Index of EU27, plotted against GDP

The most appropriate counterfactual is the closest independent country (in effective trade distance terms). We do not claim that this represents the distance that Catalonia will achieve with the rest of Spain on independence, merely that, given the data, it's the best we can do.

It's important to stress that we do not choose Portugal because is geographically close to Spain, but because it trades a lot with Spain. We do not make any claim on why Portugal is so economically close to Spain. Effective distance may be determined by geography, history, culture, or simply by chance. We do not probe the reasons for distance or proximity in this paper: we simply measure distance; although we note that distance appears to be much greater between independent countries than it does between entities which form part of the same country. Perhaps this reduced distance arises from the fact that the entities are part of the same political structure⁴. Catalans and inhabitants of the rest of Spain share many more things than Portuguese and Spanish: common regulations; greater interaction (common education, trade-fairs, contacts and networks); and they share a language. Many of these things are likely to change in with independence over the medium term. Hence we believe our proposed exercise is a natural way to approach the question of estimating the magnitude of trade costs that may arise on Catalan independence.

⁴This proposition is definitely worthy of further research! Many possible mechanism e.g. perhaps government contracting is important?

To explore possible counterfactuals we look at Spanish exports to individual countries against those countries' GDP. As can be seen on Figure 2.4, Portugal is the most positively anomalous country i.e. it is the country that trades most with Spain given its size (the vertical distance from the trend line is akin to the gravity residual in a conventional gravity equation estimation). On the same chart we also show Catalonia and the rest of Spain: it is even more anomalous.

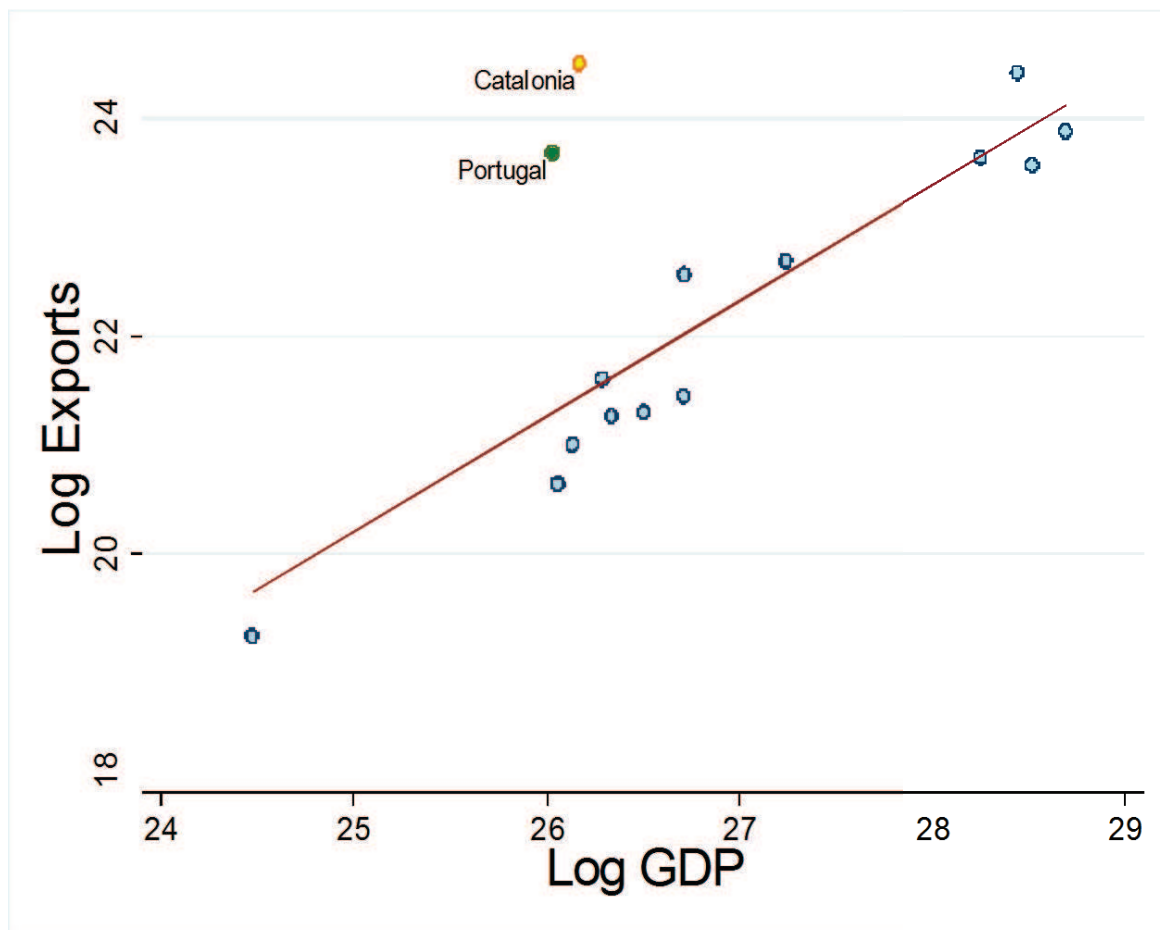


Figure 2.4: Portugal trades with Spain more than would be expected given just its GDP, based on a comparison of EU15 countries

Our choice of Portugal as the counterfactual to model Catalonian independence is therefore justified. Despite Portugal's apparent strong interactions with Spain (relative to other countries), its relationship with Spain does not look anything like Catalonia's does with the rest of Spain. Table 2.2 compares imports and exports as a percentage of GDP for Portugal & Spain against Catalonia & the rest of Spain. X_{hj} denotes exports from h to j , Y_h is GDP of h , and in each case Catalonia or Portugal is j , while Spain or the rest of Spain is h . Therefore the first row in the table is the bilateral trade as a percentage of the combined GDP of the trading partners. The second

row is the bilateral trade as a percentage of Portugal and Catalonia's GDP respectively. The third row is bilateral trade as a percentage of the Portugal and Catalonia's trade with the rest of the world. We see that Catalonia trades much much more with the rest of Spain than Portugal does with its closest trading partner, Spain.

	Portugal/Spain	Catalonia/RoS
$\frac{X_{hj}+X_{jh}}{Y_j+Y_h}$	2.51%	11.35%
$\frac{X_{hj}+X_{jh}}{Y_j}$	17.31%	60.69%
$\frac{X_{hj}+X_{jh}}{X_{Rj}+X_{jR}}$	36.27%	91.31%

3 Model

In this section we present a version of the standard Hopenhayn-Melitz model of firm heterogeneity and international trade. We model the world as consisting of three economies h, j and rest of the world R . These countries use a common currency (i.e. the nominal exchange rate is 1), but the purchasing power of this common currency can be different in the different countries. In each economy (h, j and R) there are Dixit-Stiglitz consumers. Thus, the demand for any good i in country j , is given by the demand function:

$$q_i = \left(\frac{p_i}{P_j} \right)^{-\theta} \left(\frac{Y_j}{P_j} \right)$$

- where
- $\theta > 1$ is the elasticity of substitution across goods
 - p_i is the nominal price of good i
 - P_j is the price aggregator for country j
 - q_i are the units of good i sold in country j
 - Y_j is the nominal GDP in country j

3.1 Firms

Consider an individual, monopolistically competitive, firm in economy h which takes the demand for its goods in market j as given. There is a fixed cost for creating a firm, and existing firms pay a fixed cost per period to operate in a market that is linear in the size of the market. The operating profit of being active in a country depends positively on that country's demand, and depends negatively on the economic distance between the countries of production and sales. A lower distance has positive effects at the macro level because: it increases the number of firms serving a market; and it improves the quality of the firms that serve in the country, as the more

productive firms increase labour demand in order to export. This increases wages, and drives unproductive firms out of the market.

The life cycle of a firm consists of the following stages:

- A putative firm in h chooses whether to hire \tilde{c} units of labour locally, and incur the fixed cost, $\tilde{c}W_h$, to draw some productivity from a known distribution, $\phi \sim F(\cdot)$, which is common across all economies.
- If it pays that cost, it receives a productivity, ϕ .
- If this productivity is large enough it goes ahead with production in one or more markets. Otherwise, it disappears.
- Firms die exogenously with a fixed probability, $1 - \beta$, every period.

The only input is labor and production technology is constant returns to scale subject to fixed costs (to be discussed below). The productivity of a firm from h selling in j is ϕ/δ_{hj} , where ϕ is idiosyncratic to the firm and δ_{hj} is the distance between markets h and j . δ_{hj} is the most important parameter in the model: the effective trade *distance*. It reflects how much easier it is to sell into a domestic market than to sell into a foreign market; and how much more difficult it is to sell to a country which is further away than to one that is closer. δ_{hj} effectively measures the advantages than a local producer has versus a foreign producer if both have the same intrinsic quality ϕ . It is not geographic distance, though geographic distance will have something to do with it, but a much more general “economic distance”. It will be related to language differences, regulatory differences, differences in consumer tastes and preferences (as well as, presumably, many other factors). We do not try to explain where it comes from - we just use a structural model to measure it. This article is a measurement exercise in the absence of any theory as to the explanatory factors which determine this distance. If we had a *theory of distance*, we would use it. Our point is to measure distance in the context of the model, to undertake thought experiments and make comparisons.

We assume that when selling in to the domestic market, distance equals one, and that this is the lower bound of distance. We also assume that the distance from h to j equals the distance from j to h i.e. $\delta_{hh} = \delta_{jj} = 1$ and $\delta_{hj} = \delta_{jh} \geq 1$. Nominal operating profits for a firm from h selling into market j are:

$$\tilde{\pi}_j^h = \max_{q_i} \left[p_i q_i - W_h \frac{\delta_{hj}}{\phi} q_i \right] = \max_q \left[P_j q^{1-\frac{1}{\theta}} \left(\frac{Y_j}{P_j} \right)^{\frac{1}{\theta}} - W_h \frac{\delta_{hj}}{\phi} q \right]$$

All production occurs in the firm’s local labour market (h) and so the firm hires local labour at a nominal wage rate of W_h per unit of effective labour. Profit maximisation yields revenues, labour

demand (for production) and operating profit respectively:

$$\begin{aligned}
r_j^h &= \theta \Theta \left(\frac{\phi}{\delta_{hj}} \right)^{\theta-1} \left(\frac{W_h}{P_j} \right)^{-(\theta-1)} Y_j \\
\hat{L}_j^h &= (\theta - 1) \Theta \left(\frac{\phi}{\delta_{hj}} \right)^{\theta-1} \left(\frac{W_h}{P_j} \right)^{-\theta} \frac{Y_j}{P_j} \\
\tilde{\pi}_j^h &= \Theta \left(\frac{\phi}{\delta_{hj}} \right)^{\theta-1} \left(\frac{W_h}{P_j} \right)^{-(\theta-1)} Y_j \\
\text{where } \Theta &= \frac{(\theta - 1)^{\theta-1}}{\theta^\theta}
\end{aligned}$$

The fixed costs, per unit time, of running a firm in h that sells in market j , is the cost of hiring a labour force of fixed size to deal with the expenses associated with access to this market. These expenses depend on both the size of the market that it is going to be served (Y_j/P_j) and the distance of the firm from the market (δ_{hj}). The larger the market, the more complex it is to sell there (and the larger the reward). The more distant the market is, the more complicated it is to sell there. The number of workers necessary to deal with these access expenses is assumed to be:

$$f_{hj} = c \times \delta_{hj} \times \frac{Y_j}{P_j} \quad (\text{cost parameter, } c \text{ is equal in all countries})$$

Thus, the per period net profit and labour demand for a firm in h , from selling to market j is:

$$\begin{aligned}
\pi_j^h &= \tilde{\pi}_j^h - W_h f_{hj} \\
&= \left[\Theta \left(\frac{\phi}{\delta_{hj}} \right)^{\theta-1} \left(\frac{W_h}{P_j} \right)^{-\theta} - c \delta_{hj} \right] W_h \frac{Y_j}{P_j}
\end{aligned} \tag{1}$$

$$\begin{aligned}
L_j^h &= \hat{L}_j^h + f_{hj} \\
&= (\theta - 1) \Theta \left(\frac{\phi}{\delta_{hj}} \right)^{\theta-1} \left(\frac{W_h}{P_j} \right)^{-\theta} \frac{Y_j}{P_j} + c \delta_{hj} \frac{Y_j}{P_j}
\end{aligned}$$

Profits and labour demand are increasing in productivity, ϕ and market size, Y_j . A firm from h will choose to operate in j only if its operating profit from that market exceeds its fixed cost for that market. That is, only if its productivity is high enough or the distance low enough:

$$\pi_j^h \geq 0 \iff \phi > \Phi_j^h = \left(\frac{c}{\Theta} \right)^{\frac{1}{\theta-1}} \left(\delta_{hj} \frac{P_h}{P_j} \frac{W_h}{P_h} \right)^{\frac{\theta}{\theta-1}}$$

Φ_j^h is the threshold of quality of a firm from h to operate in j . In the model we only observe firms

from h exporting to j if they have an intrinsic productivity larger than Φ_j^h . Φ_j^h will be larger (and hence we will observe a higher average quality for exporting firms) if: the distance between h and j is larger, as this increases the complexity of selling into j ; labour in h is more expensive, as this makes production more costly; the real exchange rate P_h/P_j is larger⁵, as it is less attractive to sell to j instead of h (if P_j is low relative to P_h , then the price of your good in country j will necessarily be low - otherwise you do not sell much in j). Notice that the threshold Φ_j^h is independent of the size of both markets, in particular it is independent of the size of j . This because of our assumption that the fixed costs are linear in market size. This assumption simplifies the analysis enormously, but further, we believe that it is the correct assumption given the purpose of our exercise. The assumption of fixed costs linear in size ensures that the relative size of two economies is irrelevant if their δ equals one. If this were not the case then there would be huge implications for the effect of size upon economic activity. For instance if fixed costs were independent of size (as it is normally assumed) then larger economies would be much better off - nobody would want to trade with small economies, and replicating the fact that small economies are typically more open than large ones would distort the calibrated values of distance. The average distance between h and all of its trading partners would be inversely related to the size of h , because of the need to compensate for the independence of the fixed cost and country size. Given the purpose of our exercise, this seems like the wrong assumption to make: allowing distance to be independent of size seems more appropriate, and this requires fixed costs be linear in size.

We assume (and we will check in the resulting calibrations) that $\Phi_h^h < \Phi_j^h < \Phi_R^h$. Therefore Φ_h^h acts as an existence threshold level of productivity⁶ i.e. those new firms who draw a productivity $\phi < \Phi_h^h$ choose not to produce anything. Notice further that those relatively low quality firms that do choose existence (i.e. $\phi > \Phi_h^h$) may still make a realised loss, because their positive profits may not be sufficient to cover their sunk creation costs. The realised value of creating a firm in country h is given by:

$$\begin{aligned} V^h &= \sum_{t=0}^{\infty} \beta^t (\pi_h^h(\phi) + \pi_j^h(\phi) + \pi_R^h(\phi)) - W_h \tilde{c} \\ &= \frac{\pi_h^h(\phi) + \pi_j^h(\phi) + \pi_R^h(\phi)}{1 - \beta} - W_h \tilde{c} \end{aligned}$$

Where $\pi_j^h(\phi)$ is given by equation (1) for $\phi \geq \Phi_j^h$ and by zero for $\phi < \Phi_j^h$. Notice that, since the

⁵ P_h/P_j is the real exchange rate of the goods sold in j in terms of goods sold in h . Given that the marginal utility of money in h and j are respectively $1/P_h$ and $1/P_j$, then P_h/P_j is the relative value of money in j with respect to h .

⁶Strictly this threshold productivity level is $\min\{\Phi_h^h, \Phi_j^h, \Phi_R^h\}$ so general existence conditions are complex. We short-circuit this complexity by assuming $\Phi_h^h < \Phi_j^h < \Phi_R^h$ and checking that our assumption holds in the resulting calibration.

lowest threshold of activity is to operate in the domestic market, all exporters also sell domestically, but not vice versa.

3.2 The average firm.

We assume that the firm quality distribution function is a Pareto distribution with exponent k and lowest value b . For simplicity we define:

$$\mu = k \frac{\theta}{\theta - 1}$$

We can therefore evaluate the distribution function for productivity and hence the probability of a firm, having received its productivity draw, choosing to produce:

$$\begin{aligned} Pr(\phi \geq \Phi_h^h) &= 1 - F(\Phi_h^h) = \left(\frac{b}{\Phi_h^h} \right)^k \\ &= b^k \left(\frac{\Theta}{c} \right)^{-\frac{\mu}{\theta}} \left(\frac{W_h}{P_h} \right)^{-\mu} \end{aligned}$$

We assume $k > \theta - 1$ so that average profits, revenues and labour demand are defined⁷.

The average realised profit of the observed firms is the expected profit of a firm, conditional on existence:

$$\begin{aligned} \bar{\pi}^h &= \int_b^\infty [\pi_h^h(\phi) + \pi_j^h(\phi) + \pi_R^h(\phi)] \frac{dF(\phi)}{1 - F(\Phi_h^h)} \\ &= cW_h \left[\frac{1}{\theta} - \frac{1}{\mu} \right]^{-1} \frac{1}{\mu} \left\{ \frac{Y_h}{P_h} + \left(\frac{P_j}{P_h} \right)^\mu (\delta_{hj})^{1-\mu} \frac{Y_j}{P_j} + \left(\frac{P_R}{P_h} \right)^\mu (\delta_{hR})^{1-\mu} \frac{Y_R}{P_R} \right\} \\ &= cW_h \left[\frac{1}{\theta} - \frac{1}{\mu} \right]^{-1} \frac{1}{\mu} \frac{D_h}{P_h} \end{aligned}$$

Where we define effective (nominal) demand in economy h as:

$$D_h = P_h \left\{ \frac{Y_h}{P_h} + \left(\frac{P_j}{P_h} \right)^\mu (\delta_{hj})^{1-\mu} \frac{Y_j}{P_j} + \left(\frac{P_R}{P_h} \right)^\mu (\delta_{hR})^{1-\mu} \frac{Y_R}{P_R} \right\} \quad (2)$$

The average labour demand⁸ is the sum of the labor demand used for selling into each of the three markets, and average revenues are the sum of the revenues from selling into each of the three

⁷Profits, revenues and labour demand also have a Pareto distribution, but with exponent $k + 2 - \theta$. The means of these Pareto distributions are only defined if $k + 2 - \theta > 1$ i.e. $k > \theta - 1$.

⁸Notice that labour demand from the average firm does not depend on the wage. The total labour demand it is going to depend on wages though because the number of firms will depend on wage.

markets:

$$\begin{aligned}\bar{L}^h &= c \left[\frac{1}{\theta} - \frac{1}{\mu} \right]^{-1} \left(1 - \frac{1}{\mu} \right) \frac{D_h}{P_h} \\ \bar{r}^h &= cW_h \left[\frac{1}{\theta} - \frac{1}{\mu} \right]^{-1} \frac{D_h}{P_h}\end{aligned}$$

3.3 General Equilibrium

The world economy in this model will be in steady state general equilibrium if we have steady state equilibrium in three markets:

1. Financial Markets:

Within each country there are perfect financial markets which allow prospective firms to borrow to finance firm creation via contingent contracts which, in equilibrium, will be repaid using any realised profits. Free entry for entrepreneurs to create firms therefore means that the expected value of firm creation will be driven to zero:

$$\begin{aligned}E[V^h] &= F(\Phi_h^h) \times 0 + (1 - F(\Phi_h^h)) \times \frac{\bar{\pi}^h}{1 - \beta} - W_h \tilde{c} = 0 \\ \text{i.e. } \left(\frac{W_h}{P_h} \right)^{-\mu} \frac{D_h}{P_h} &= \frac{\tilde{c}(1 - \beta)}{cb^k} \left(\frac{c}{\Theta} \right)^{\frac{\mu}{\theta}} \left(\frac{1}{\theta} - \frac{1}{\mu} \right) \mu\end{aligned}\quad (3)$$

2. Labour Markets:

Let the number of firms producing in economy h be M_h , then we know that in each period $(1 - \beta)M_h$ firms die. In steady state, M_h is constant and so the number of entrepreneurs who hire labour in an attempt to create a firm must be such that the resulting number of firms that choose to operate (i.e. who have $\phi > \Phi_h^h$) is equal to $(1 - \beta)M_h$ i.e. the labour employed in paying the fixed creation cost:

$$L_{creation} = \frac{1 - \beta}{1 - F(\Phi_h^h)} M_h \tilde{c} = cM_h \left(\frac{D_h}{P_h} \right) \left(\frac{1}{\theta} - \frac{1}{\mu} \right)^{-1} \frac{1}{\mu}$$

Total demand for effective labour is the sum of $L_{creation}$ and the labour employed by the firms that have decided to go ahead ($M_h \bar{L}^h$). Effective labour supply, S_h is an exogenous parameter, different in different economies, that we will calibrate to. This parameter can be interpreted as the population in a country (N_h from data) multiplied by an intrinsic, unobserved, productivity (A_h) which can be calculated given our calibrated value of S_h .

Assuming effective labour supply equals demand for effective labour:

$$\left(\frac{1}{\theta} - \frac{1}{\mu}\right) S_h = cM_h \frac{D_h}{P_h} \quad (4)$$

3. Goods Markets:

In this exercise, we allow for a fiscal transfer from j to h (remember the existence of a fiscal transfer from Catalonia to the rest of Spain provides some of the momentum towards independence) of FS_jW_j . F is the percentage of GDP transferred from j to h (which in principle could be negative). There is no fiscal transfer to or from R .

$$Y_j = (1 - F)S_jW_j \quad (5)$$

$$Y_h = S_hW_h + FS_jW_j \quad (6)$$

$$Y_R = S_RW_R \quad (7)$$

To close the model we equalise income and aggregate demand by imposing a balance of payments for each economy. Payments are balanced, albeit not necessarily bilaterally balanced. Thus, in each country total export earnings plus any net fiscal transfers received are equal to total import expenditures. There can be bilateral surpluses or deficits, but overall there is a balance of payments:

$$FS_jW_j + X_j^h + X_R^h = X_h^j + X_h^R \quad \text{i.e. BoP in } h$$

$$X_h^j + X_R^j - FS_jW_j = X_h^h + X_j^R \quad \text{i.e. BoP in } j$$

If h and j are in balance, R must be too as the above conditions imply:

$$X_h^R + X_j^R = X_h^h + X_j^j \quad \text{i.e. BoP in } R$$

Exports from h to j are:

$$\begin{aligned} X_j^h &= M_h \frac{1 - F(\Phi_j^h)}{1 - F(\Phi_h^h)} \int_{\Phi_j^h} r_j^h(\phi) \frac{dF(\phi)}{1 - F(\Phi_j^h)} \\ &= cM_h \frac{W_h}{P_h} \left(\frac{P_h}{P_j} \delta_{hj}\right)^{1-\mu} Y_j \left(\frac{1}{\theta} - \frac{1}{\mu}\right)^{-1} \end{aligned}$$

Balance of payments in h :

$$\begin{aligned} & \frac{FS_j W_j}{c} \left(\frac{1}{\theta} - \frac{1}{\mu} \right) + M_h \frac{W_h}{P_h} \left\{ \left(\frac{P_h}{P_j} \delta_{hj} \right)^{1-\mu} Y_j + \left(\frac{P_h}{P_R} \delta_{hR} \right)^{1-\mu} Y_R \right\} \\ & = Y_h \left\{ M_j \frac{W_j}{P_j} \left(\frac{P_j}{P_h} \delta_{hj} \right)^{1-\mu} + M_R \frac{W_R}{P_R} \left(\frac{P_R}{P_h} \delta_{hR} \right)^{1-\mu} \right\} \end{aligned} \quad (8)$$

Balance of payments in j :

$$\begin{aligned} & M_j \frac{W_j}{P_j} \left\{ \left(\frac{P_j}{P_h} \delta_{hj} \right)^{1-\mu} Y_h + \left(\frac{P_j}{P_R} \delta_{jR} \right)^{1-\mu} Y_R \right\} - \frac{FS_j W_j}{c} \left(\frac{1}{\theta} - \frac{1}{\mu} \right) \\ & = Y_j \left\{ M_h \frac{W_h}{P_h} \left(\frac{P_h}{P_j} \delta_{hj} \right)^{1-\mu} + M_R \frac{W_R}{P_R} \left(\frac{P_R}{P_j} \delta_{jR} \right)^{1-\mu} \right\} \end{aligned} \quad (9)$$

3.4 Model Solution

Equations (2), (3) & (4) (all $\times 3$ for h, j & R), plus equations (5), (6), (7), (8), & (9) constitute 14 equations in 15 unknowns ($\{D_h, D_j, D_R, P_h, P_j, P_R, Y_h, Y_j, Y_R, W_h, W_j, W_R, M_h, M_j, M_R\}$) which can be solved by assuming that one of the price indices is normalised to 1. The system has 12 parameters ($\{\theta, \mu, \delta_{hj}, \delta_{hR}, \delta_{jR}, c, \tilde{c}, b, \beta, S_h, S_j, S_R\}$). We can simplify the system greatly by making the transformations detailed in Appendix 1. This produces a system of 10 equations (see below) in 10 endogenous variables ($\{\tilde{d}_h, \tilde{d}_j, Q_{hj}, \tilde{Q}_{hR}, \tilde{y}_h, \tilde{y}_j, \tilde{w}_h, \tilde{w}_j, \tilde{M}_h, \tilde{M}_j\}$) with 6 parameters ($\{\theta, \mu, \delta, \Delta_h, \Delta_j, s_j\}$). s_j is the relative effective size of j with respect to h . This is equal to $(N_j A_j)/(N_h A_h)$. Given that we take N_j and N_h directly from the data, s_j gives us the implied ratio of worker's productivity. Notice however that this does not map directly into observed productivity because it is filtered by the distribution of firm's qualities. Δ_h and Δ_j are measures of effective size of the rest of the world from the viewpoint of h and j respectively. This is the effective size filtered by the economic distance between the rest of the world and h & j respectively. The larger it is, the more economic interaction there is with R . The ratio of Δ_h to Δ_j is a measure of the relative distance to the rest of the world of h with respect to j .

The equations of the model are:

$$\begin{aligned}
0 &= 1 - \frac{\tilde{y}_h + \delta^{1-\mu} Q_{hj}^{-\mu} \tilde{y}_j + \Delta_h \tilde{Q}_{hR}^{-\mu}}{\tilde{d}_h} \\
0 &= 1 - \frac{\tilde{y}_j + \delta^{1-\mu} Q_{hj}^{\mu} \tilde{y}_h + \Delta_j \tilde{Q}_{hR}^{-\mu} Q_{hj}^{\mu}}{\tilde{d}_j} \\
0 &= 1 - \left(\frac{1}{\theta} - \frac{1}{\mu} \right) \mu \Theta^{-\frac{\mu}{\theta}} \tilde{w}_h^{\mu} / \tilde{d}_h \\
0 &= 1 - \left(\frac{1}{\theta} - \frac{1}{\mu} \right) \mu \Theta^{-\frac{\mu}{\theta}} \tilde{w}_j^{\mu} / \tilde{d}_j \\
0 &= 1 - \tilde{d}_h \tilde{M}_h \left(\frac{1}{\theta} - \frac{1}{\mu} \right)^{-1} \\
0 &= 1 - \tilde{d}_j \tilde{M}_j \frac{1}{s_j} \left(\frac{1}{\theta} - \frac{1}{\mu} \right)^{-1} \\
0 &= 1 - \frac{\tilde{w}_h Q_{hj} + F s_j \tilde{w}_j}{\tilde{y}_h Q_{hj}} \\
0 &= 1 - \frac{(1-F) s_j \tilde{w}_j}{\tilde{y}_j} \\
0 &= 1 - \frac{\tilde{y}_h \left(\tilde{M}_j \tilde{w}_j Q_{hj}^{\mu-1} \delta^{1-\mu} + \tilde{Q}_{hR}^{\mu-1} \Delta_h \right)}{F \tilde{M}_j \tilde{w}_j \tilde{d}_j \frac{1}{Q_{hj}} + \tilde{M}_h \tilde{w}_h \left(Q_{hj}^{-\mu} \delta^{1-\mu} \tilde{y}_j + \tilde{Q}_{hR}^{-\mu} \Delta_h \right)} \\
0 &= 1 - \frac{\tilde{y}_j \left(\tilde{M}_h \tilde{w}_h Q_{hj}^{1-\mu} \delta^{1-\mu} + \left(\frac{Q_{hj}}{\tilde{Q}_{hR}} \right)^{1-\mu} \Delta_j \right)}{\tilde{M}_j \tilde{w}_j \left(Q_{hj}^{\mu} \delta^{1-\mu} \tilde{y}_h + \left(\frac{\tilde{Q}_{hR}}{Q_{hj}} \right)^{-\mu} \Delta_j - F \tilde{d}_j \right)}
\end{aligned}$$

For a given set of parameters, we determine equilibrium by applying the Newton-Raphson method to find the roots of the above system of equations. We apply the Simulated Method of Moments (see Appendix 1) until we equate a vector of data targets with their modelled equivalents.

4 Data

All data is for the year 2005.

4.1 Aggregate Data

The aggregate data is data on GDP and on trade flows (both goods and services) at the national level and at the level of Spain's autonomous communities. Appendix 2 details exactly what the data is and where it comes from.

4.2 Micro Data

The microdata, Encuesta Sobre Estrategias Empresariales (ESSE), is data on the distribution of Spanish firms. Appendix 2 details what the data is and where it comes from. From this data, we identify firms as Catalan or as coming from the rest of Spain. We observe their total sales and in which region (local, national or international) these sales were made.

4.3 Selection of k and θ

The other *data* that we use is the economic literature. We take the elasticity of substitution, θ , from the literature and follow the procedure that others have followed in determining the pareto distribution parameter, k . Bernard, Eaton, Jensen, and Kortum (2003) (BEJK) select a θ of 3.79 to match the size and productivity advantage of US firms that export⁹. Many papers use $\theta = 3.8$ following BEJK, see e.g. Ghironi and Melitz (2005), Davis and Harrigan (2011), and Bernard, Redding, and Schott (2007).

Some papers calibrate k to match the standard deviation of log domestic sales in the US (as found by BEJK). See e.g. Davis and Harrigan (2011) ($k = 3.4$), Ghironi and Melitz (2005) ($k = 3.4$), Demidova (2008) ($k = 3.3$), Felbermayr and Jung (2012) ($k = 3.3$). The Standard deviation of log firm sales in the model of Section 3 is given by $(\theta - 1)/k$. Therefore, using the microdata, and conditional on $\theta = 3.8$:

	Stdev log sales	Implied k
Observed US data (BEJK 2003)	1.67	1.7
Simulated US data (BEJK 2003)	0.84	3.3
Spain data (ESSE survey)	1.90	1.5

For the mean firm profits, revenue and labour demand to be defined, a parameter restriction of $\theta - 1 < k$ is imposed i.e. the value of k must be greater than $k_{\min} = 2.8$. It is impossible to simultaneously match a standard deviation of log sales, which in our model is given by $(\theta - 1)/k$, with a value greater than 1 as well as imposing the $\theta - 1 < k$ restriction from the model.

5 Results

To generate our results we calibrate the model in two configurations in which the parameters not calibrated are:

⁹Though in BEJK markups (and not productivity) are drawn from a pareto distribution, so the shape parameter used in their paper is not applicable.

	Configuration 1	Configuration 2
θ	3.8	3.8
k	3.3	2.8001
μ	4.5	3.8

Configuration 1 uses the parameters of BEJK 2003 for a simulated distribution of firms in the US, while Configuration 2 is an attempt to use the Spanish firm data. The standard deviation of firm size in Spain appears to be substantially larger than in the US and in order to match this moment we would need to have a very low value of k , which would be incompatible with our model given the established value of θ . Because of this we choose the minimum value for k compatible with our restrictions, 2.8. The effect of using a larger k than suggested by the data is to underestimate the cost of larger distance (with respect to this choice of k)

We take the transfer of Catalonia to rest of Spain to be 6.5% of the GDP of Catalonia. This is the official number for the transfer with the methodology of "flujo beneficios" for the year 2005. In 2009 this number would be 5.8% of the GDP of Catalonia. We do not allow for an international transfer (e.g. EU structural funds) in the Portugal - Spain calibration.

The parameters that we calibrate using SMM are therefore $\{s_j, \delta, \Delta_h, \Delta_j\}$ (where h labels Spain or the rest of Spain, and j labels Portugal or Catalonia respectively in each calibration). To calibrate these parameters we target:

- The interaction between h and j : $(X_{hj} + X_{jh})/(Y_h + Y_j)$
- Total trade in h : $(X_{hj} + X_{jh} + X_{hR} + X_{Rh})/Y_h$
- Total trade in j : $(X_{hj} + X_{jh} + X_{jR} + X_{Rj})/Y_j$
- Relative GDP: Y_h/Y_j

To validate the model results we use the average sales of Catalan Firms in the rest of Spain, divided by the average international sales of Catalan firms. It is not clear whether this ratio will be less than or greater than 1: the numerator includes firms with lower productivity and is the average of sales into a smaller market, however the denominator is the average of sales into a more distant market. *Cat* firm size ratio is defined as:

$$Cat \text{ firm size ratio} = \frac{\int_{\Phi_h^j}^{\infty} r_h^j \frac{dF(\phi)}{1-F(\Phi_h^j)}}{\int_{\Phi_R^j}^{\infty} r_R^j \frac{dF(\phi)}{1-F(\Phi_R^j)}} = \dots = \frac{\delta^{1-\mu} \tilde{y}_h}{\Delta_j \tilde{Q}_{hR}^{-\mu}}$$

5.1 Catalonia's distance with Rest of Spain

Catalonia/RoSpain	Calibrated Parameters	
	$\theta = 3.8, k = 3.3$	$\theta = 3.8, k = 2.8$
δ	1.4336	1.5643
Δ_h	0.051654	0.0015326
Δ_j	0.038254	0.00113504
s_j	0.27780	0.28306

Catalonia/RoSpain	MODEL		DATA
Targets	$\theta = 3.8, k = 3.3$	$\theta = 3.8, k = 2.8$	
$\frac{X_{hj}+X_{jh}}{Y_h+Y_j}$	0.11352	0.11352	0.11352
$\frac{X_{hj}+X_{jh}+X_{hR}+X_{Rh}}{Y_h}$	0.68769	0.68769	0.68769
$\frac{X_{hj}+X_{jh}+X_{jR}+X_{Rj}}{Y_j}$	1.27151	1.2715	1.2715
$\frac{Y_h}{Y_j}$	4.3460	4.3460	4.3460

Per-capita incomes are matched to the data, and we derive a ratio of efficiency labour per capita in Catalonia to the rest of Spain of:

Catalonia	$s_j \times \frac{N_h}{N_j}$
$\theta = 3.8, k = 3.3$	1.47
$\theta = 3.8, k = 2.8$	1.50

For model validation we look at how the distribution of firm sizes implied by the model matches the distribution in the data. It does so remarkably well:

Model Validation	MODEL	DATA
<i>Cat</i> firm size ratio	0.92482	0.88

5.2 Portugal's distance with Spain

Portugal/Spain	Calibrated Parameters	
	$\theta = 3.8, k = 3.3$	$\theta = 3.8, k = 2.8$
δ	2.3481	2.8876
Δ_h	0.047784	0.0014158
Δ_j	0.017896	0.00053021
s_j	0.20548	0.21263

Per-capita incomes are matched to the data, and we derive a ratio of efficiency labour per capita in Portugal to Spain of:

Portugal	$s_j \times \frac{N_h}{N_j}$
$\theta = 3.8, k = 3.3$	0.91
$\theta = 3.8, k = 2.8$	0.93

Notice that the distance implied between Portugal and Spain is substantially larger than the distance implied between Catalonia and the Rest of Spain.

Portugal/Spain	MODEL		DATA
	$\theta = 3.8, k = 3.3$	$\theta = 3.8, k = 2.8$	
Targets			
$\frac{X_{hj}+X_{jh}}{Y_h+Y_j}$	0.025115	0.025115	0.025115
$\frac{X_{hj}+X_{jh}+X_{hR}+X_{Rh}}{Y_h}$	0.56986	0.56986	0.56986
$\frac{X_{hj}+X_{jh}+X_{jR}+X_{Rj}}{Y_j}$	0.65051	0.65051	0.65051
$\frac{Y_h}{Y_j}$	5.8942	5.8942	5.8942

5.3 Experiment 1: Changing Distance.

$\theta = 3.8, k = 3.3$	$\delta = 1.4336$	$\delta = 2.3481$	% change
Y_j	B\$211.5	B\$191.7	-9.4%
Y_h	B\$919.3	B\$901.2	-2.0%
$\frac{X_{hj}+X_{jh}}{Y_j}$	60.7%	13.7%	-47.0%
$\frac{X_{jR}+X_{Rj}}{Y_j}$	66.5%	80.9%	+14.4%
Φ_j^j	n/a	n/a	-12.5%
Φ_h^j	n/a	n/a	+79.7%
Φ_R^j	n/a	n/a	-6.4%
$\theta = 3.8, k = 2.8$	$\delta = 1.5643$	$\delta = 2.8876$	% change
Y_j	B\$211.5	B\$187.4	-11.4%
Y_h	B\$919.3	B\$896.9	-2.4%
$\frac{X_{hj}+X_{jh}}{Y_j}$	60.7%	13.7%	-47.0%
$\frac{X_{jR}+X_{Rj}}{Y_j}$	66.5%	81.2%	+14.7%
Φ_j^j	n/a	n/a	-15.2%
Φ_h^j	n/a	n/a	+107.7%
Φ_R^j	n/a	n/a	-7.6%

We equate nominal GDP to its value in the data, which implies a value for the parameters that were eliminated with the variable substitutions. This is irrelevant, as we only make comparisons on changes or ratios of GDP from where these parameters disappear. The increase in distance with the largest trading partner of Catalonia has dramatic effects on trade and GDP. The deadweight loss (the percentage change in $Y_h + Y_j$) is 3.3% of GDP in configuration 1, and 4.1% in configuration

2. In any case a large GDP fall. The degree of interaction with the rest of Spain becomes similar to the Portuguese one. It is actually somewhat lower due to the fact that Catalonia is closer to the rest of the World than Portugal is. The degree of interaction with the rest of the world increases, as Catalan firms find harder to sell in the rest of Spain, and thus reallocate towards the rest of the world. Notice that this increase is not of the same magnitude than the decrease in trade with the rest of Spain due to the fact that there is no decrease in the intrinsic distance with the rest of the world. Likewise Catalan consumers become less prone to consume Spanish products.

The most remarkable result reported in this table is the implications that changing the distance has on firm composition, and via this mechanism, on TFP. Wages in Catalonia fall, and this makes less productive firms viable. Notice the fall in the threshold of quality of domestic firms. Moreover, the threshold of quality for exporting to the rest of Spain rises dramatically, as it is much harder to overcome the larger distance, while the threshold to export to the rest of the world actually falls due to lower wages. There is an additional composition effect, as wages fall, less productive firms account for a larger share of employment, pushing TFP down.

5.4 Experiment 2: Changing Transfers.

$\theta = 3.8, k = 3.3$	$\delta = 1.4336, F = 0.065$	$\delta = 2.3481, F = 0$	% change
Y_j	B\$211.5	B\$206.2	-2.5%
Y_h	B\$919.3	B\$886.1	-3.6%
$\frac{X_{hj}+X_{jh}}{Y_j}$	60.7%	12.9%	-47.8%
$\frac{X_{jR}+X_{Rj}}{Y_j}$	66.5%	76.9%	+10.4%
Φ_j^j	<i>n/a</i>	<i>n/a</i>	-11.8%
Φ_h^j	<i>n/a</i>	<i>n/a</i>	+83.2%
Φ_R^j	<i>n/a</i>	<i>n/a</i>	-4.9%
$\theta = 3.8, k = 2.8$	$\delta = 1.5643, F = 0.065$	$\delta = 2.8876, F = 0$	% change
Y_j	B\$211.5	B\$201.7	-4.6%
Y_h	B\$919.3	B\$881.7	-4.1%
$\frac{X_{hj}+X_{jh}}{Y_j}$	60.7%	12.9%	-47.8%
$\frac{X_{jR}+X_{Rj}}{Y_j}$	66.5%	77.1%	+10.6%
Φ_j^j	<i>n/a</i>	<i>n/a</i>	-14.4%
Φ_h^j	<i>n/a</i>	<i>n/a</i>	+112.6%
Φ_R^j	<i>n/a</i>	<i>n/a</i>	-5.8%

Obviously money that is not transferred to the rest of Spain increases Catalan GDP. This increase is more than 6.5% of GDP (it's around 7.6%) due to multiplier effects from the IRS implied in the Dixit-Stiglitz framework. Notice though that the deadweight losses are very similar to before

(3.4% and 4.2% in each configuration). What changes here is the distribution of the losses, which now fall more heavily on the rest of Spain. This distribution of losses does not greatly affect the impacts on the firm quality distribution. The loss in Catalonia is mitigated by not paying the fiscal transfer, but these losses are still substantial, whilst the loss in the rest of Spain is now larger.

5.5 Border Effects Within The EU

We see that Catalonia is much closer to the rest of Spain than Portugal is to Spain. Can we interpret this as an effect of belonging to a single country? Are sub national borders "thinner" than the "thick" borders between nations? The single market in the EU is an attempt to create the trade benefits of a single country across Europe. Are national borders within the EU "thinner" than borders across the EU/non-EU divide? The following table shows the distances between the (rest of the) EU and Norway, Sweden, Switzerland and Austria, calibrated from 3 country models also featuring the rest of the world:

$\theta = 3.8, k = 3.3$	δ
Norway vs the EU	2.5700
Sweden vs rest of the EU	2.4220
Switzerland vs the EU	2.2991
Austria vs the rest of the EU	2.2166
Portugal vs Spain (for Comparison)	2.3481

The borders between these countries and the (rest of the) EU look similar to each other and similar to the border between Portugal and Spain¹⁰. Obviously there are many potential endogeneity issues here e.g. perhaps Switzerland did not join the EU since there were some costs, and it already had all the distance and trade benefits that it could get from the EU countries, whilst countries that could gain trade benefits on joining did so and have now reduced their distance to something akin to Switzerland's. However, these results suggest that we may be able to draw the conclusion that the EU/non EU border looks like any other country border, i.e. it is much "thicker" than the "thin", within country borders, such as that between Catalonia and the rest of Spain.

5.6 Comparing the results with the Literature

Our estimated impact on Catalonian trade is perfectly consistent with the literature on border effects taking into account endogeneity and trade with the rest of the world. In particular our results are in line with [Anderson and van Wincoop \(2003\)](#) (AvW) [quoting from section V]:

¹⁰The results are suggestive of a small border effect from the EU though.

“Based on the estimated multi-country model, international trade among ROW countries [i.e., not US and Canada] drops to a fraction 0.71 of that under free trade, while intranational trade rises on average by a factor 3.8. This implies a factor 5.4 (3.8/.71) increase in intranational trade relative to international trade.”

Catalan trade with the rest of Spain in Experiment 1 (we label this results *Virtual_Catalonia_1*, and choose Experiment 1 rather than Experiment 2 so as not to contaminate the effect of simply changing distance with the additional effect from the fiscal transfer¹¹) is about 80% lower than in the data¹². This is the same order as AvW measured for the average border effect. If we call $X(0)$ Catalonian imports and exports to the rest of Spain under *free trade* (in AvW parlance this is trade when there are no frictions at all between any country in the world), $X(1)$ imports and exports under a political union (i.e. within national borders, in our case, the trade that exists between Catalonia and the rest of Spain in the data), and $X(2)$ the imports and exports to the rest of Spain from *Virtual_Catalonia_1* (i.e. the trade between two countries with normal borders while all the other countries also have normal borders). Then AvW estimate:

$$\frac{X(1)}{X(0)} = 3.8 \quad \& \quad \frac{X(2)}{X(0)} = 0.71 \quad \Rightarrow \quad \frac{X(2)}{X(1)} = 0.19$$

Thus, the AvW analysis suggests that we should expect to see a trade fall of around 80% when comparing trade across a border to trade without the border (but with the rest of the world having such a border). Our result is consistent with AvW, and so it appears that Catalonia is not exceptional when viewed under the light of the border effect. The reason why it trades “so much” and why it concentrates trade in Spain is just the “thick” border with the rest of the world and the “thin” border (small distance) with the rest of Spain.

5.7 The Gains from Trade

The gains from trade in the literature are usually expressed as a welfare cost on moving to an autarkic state. Arkolakis, Costinot, and Rodríguez-Clare (2012) (ACRC) develop a standard formula for the gains from trade, based only upon the share, λ , of expenditure on domestic goods (observable from data), and the elasticity, ϵ , of trade flows with respect to variable trade costs (from gravity equation estimation), for a whole class of models i.e. essentially all models that have a gravity equation representation have the same mapping from trade flows to welfare gains/losses.

¹¹i.e. *Virtual_Catalonia_1* is Catalonia as distant from the rest of Spain as Portugal is from Spain, but still paying the fiscal transfer.

¹²(13.7% × \$191.7)/(60.7% × \$211.5) = 20% under Configuration 1.

From ACRC, the gravity equation needed to estimate ϵ is:

$$\ln X_j^h = A_h + B_j + \epsilon \ln \delta_{hj} + \nu_j^h$$

where A_h & B_j are a sum of regressors multiplied by their coefficients
 δ_{hj} can be identified with the trade distance we discuss
and ν is the residual, zero for a theoretical model as opposed to an empirical exercise.

The model of this paper is within the class of models described by ACRC. The gravity equation implied our model is:

$$\log X_j^h = \log Y_h - \log D_h + \log Y_j + (1 - \mu) \log Q_{hj} + (1 - \mu) \log \delta_{hj} + 0$$

Clearly the model presented in this paper has $\epsilon \equiv 1 - \mu$. We can therefore calculate the gains from trade both using ACRC formula and by setting distances to infinity and running the model. They agree as required. The ACRC formula is: Cost of autarky = $\lambda^{1/\epsilon} - 1$, and the calculated costs are:

$k = 3.3; \theta = 3.8$	λ	ACRC Formula Result	Modelled Result
Spain	71.5%	10.1%	10.1%
Portugal	67.5%	12.0%	12.0%
Catalonia (data)	39.9%	N/A	N/A
Catalonia (no transfer)	39.4%	30.7%	30.7%
Virtual_Catalonia_2	55.1%	18.7%	18.7%

Note that the model is not capable of running with a fiscal transfer from Catalonia to the rest of Spain under autarky (and the ACRC formula is not valid in the presence of such a transfer). It is impossible to achieve a balance of payments if there is a fiscal transfer in one direction that is not offset by trade in the other direction. It is for this reason that the table shows *Virtual_Catalonia_2*: Catalonia as distant from the rest of Spain as Portugal is from Spain, and with no fiscal transfer¹³.

The trade literature estimates very low gains from trade (see Table IX of Eaton and Kortum (2002) for a set of much lower estimates of the costs of autarky than those above) and the results presented here are somewhat anomalous. The discrepancy arises from the estimation of the elasticity, ϵ , of imports with respect to variable trade costs. Most of the theoretical trade literature finds data on the variable trade costs and run regressions to determine this elasticity. This leaves the possibility that these trade costs are underestimated and so the regression coefficient is

¹³If GDP Catalonia = 100%, then GDP Virtual Catalonia 1 = 90.6%, GDP Catalonia (No Transfer) = 107.4%, GDP Virtual Catalonia 2 = 97.5%, and GDP Catalonia Autarky = 82.1%. The ACRC Cost of Autarky calculation is then of the form: $107.4\%/82.1\% = 130.7\%$ & $97.5\%/82.1\% = 118.7\%$.

overestimated. Indeed the [Anderson and van Wincoop \(2004\)](#) survey finds gravity equation based estimates for this trade elasticity in the range $(-10, -5)$ as compared with the values of -3.5 and -2.8 that are used here (in Configuration 1 and 2 respectively). The alternative methodology used in this paper (whereby we use: accepted primitive parameters for elasticity of substitution between goods; then try to match moments of the firm distribution; and then use a structural model to calculate the elasticity of trade flows with respect to variable trade costs) perhaps provides a clue to a puzzle in the trade literature. ACRC note in their concluding remarks that [quoting from section 6]:

“If many of our theoretical models predict the same gains from trade, how can these gains be so much smaller than the reduced-form estimates uncovered by empirical researchers? For instance, [Feyrer \(2009\)](#) concludes that an increase in trade volumes of 10% implies an increase in real income of 5%. While this elasticity lies below the previous estimates of [Frankel and Romer \(1999\)](#), it is an order of magnitude larger than the elasticity implied by gravity models. How does one reconcile theory and empirics?”

The methodology used in this exercise suggests that part of the discrepancy between theory and empirics may be due to the fact that existing theoretical trade literature takes the measure of distance from data, and with it they estimate the elasticity of trade to distance. But this data on distance and frictions is a composite of tariff and fiscal distance which may omit important frictions. We calibrate frictions - we do not get them from data. Our frictions are larger than the frictions typically used, and thus, our gains from trade are larger too.

There is one important feature to note about the gains from trade: economically integrated regions within a country have almost as much to lose from losing this close integration and becoming a normal independent country, as normal independent countries have from becoming autarkic. This is consistent with the large positive welfare effects of moving to a *zero gravity world* in [Eaton and Kortum \(2002\)](#) (see Table X). We can show this graphically by generating a series of welfare results for Catalonia, changing only distance with the rest of Spain between each result and graphing on a single plot, see Figure 5.1. Independent countries exist along the ‘flat part’ of the ‘distance vs GDP’ curve, whereas sub-national regions exist in the steeper region where gains from trade are larger. This suggests that there may exist large economic benefits, through closer trade links, of further integration at the EU level (since we observe that the borders between EU member states are of the “thick”, independent country type).

Catalonia’s cost of autarky figure is very high, but this is because it is the composition of two losses. The first loss is the increase in frictions with the rest of Spain, that were very low (much lower than the normal between countries), and are then those of a normal country. The second loss is the increase in frictions from having normal borders with the rest of Spain and with the

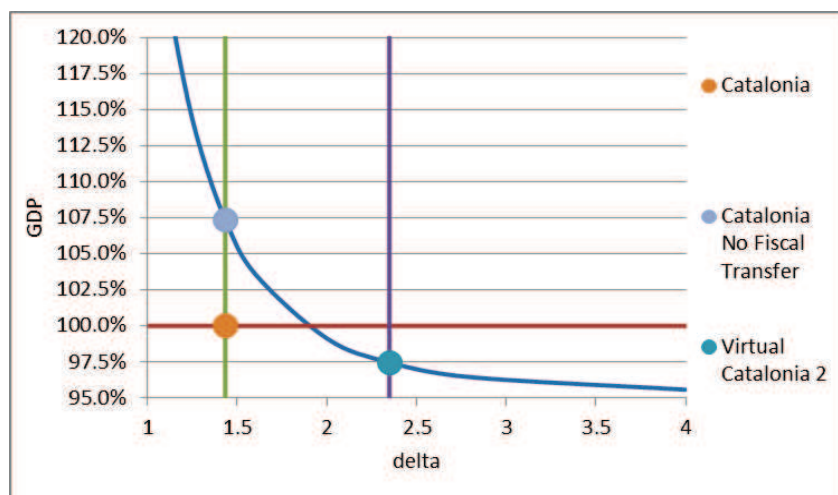


Figure 5.1: Catalan GDP as a function of effective distance with the rest of Spain. Note that the limit of infinite distance on this plot does not correspond to the autarkic Catalonia - the only distance that is being changed in this plot is the distance with the rest of Spain, the distance with the rest of the world is unaltered.

rest of the world, to having infinite frictions. This composite effect means that we would expect the cost of autarky for sub-national regions to be much higher than is estimated for independent countries of equivalent size. Therefore the costs, given the current institutional framework, of the break-up of EU member states into smaller states within the EU, may be relatively high.

6 Conclusions

In this article, we have shown evidence that Catalonia seems to have an exceptionally high level of economic integration with the rest of Spain. The level of this integration is such that Catalonia does not look like other independent countries, and we suggest that it is unreasonable to believe that this level of economic integration will persist in the event of independence. A more reasonable assumption is that Catalonia will become a country which is relatively close to the rest of Spain, but not exceptionally close compared with other independent countries. The natural counterfactual to use to model an independent Catalonia is Portugal, which is economically close to Spain, but not exceptionally close like Catalonia currently is.

The exercise that we undertake is then to calibrate our structural model, based on [Melitz \(2003\)](#), to Catalonia, the rest of Spain, and the rest of the world; and to Portugal, Spain, and the rest of the world. These calibrations produce a set of parameters which, when plugged into the model, reproduce the incomes and trade flows seen in the data. In the calibrations, Portugal is much further from Spain than Catalonia is from the rest of Spain. The policy experiment undertaken is

simply to replace Catalonia's distance with the rest of Spain with Portugal's distance from Spain, and to observe the impacts upon incomes and trade flows. We show that the losses associated with this increase in distance from the rest of Spain are large. The combined GDP loss of increasing the distance is of between 3.3 and 4.2 percent. From the point of view of Catalonia, the loss if unaccompanied by a fiscal gain is huge: of the between 9.4 and 11.4 percent. If Catalonia gains from not paying a fiscal transfer to the rest of Spain then the loss for Catalonia is between 2.5 and 4.6 percent: smaller, but still substantial. Clearly in this case the loss for the rest of Spain would be much larger.

Given the form of the structural model used to model this, a large share of the GDP loss on independence is a consequence of a worsening in the distribution of firm quality in Catalonia: mediocre firms that would find it impossible to survive when facing direct competition from the rest of Spain, would find it profitable to survive in the independence scenario. This is because they are sheltered from more intense competition by the larger distance. In the Melitz model, there are two dimensions that characterise a firm: its quality and its location. When distance increases location becomes more salient. Unproductive but local firms find themselves in a better competitive position, and they access more of what is now a captive market. This reallocation towards inefficiency is a plausible mechanism for generating costs of independence: the rise of mediocrity.

This paper makes a methodological contribution in calibrating economic distances given observed trade flows within a structural model that has been calibrated to accepted primitive parameters and to firm distribution statistics. The resulting costs of autarky are very large relative to the economic literature, a discrepancy which arises from two sources. The headline reason for this discrepancy is that in the case of sub-national entities, the cost of autarky is magnified due to the exceptionally high degree of integration that exists within countries. There is a genuinely higher cost of autarky due to this effect. There is an additional component which arises from the methodology: primitive parameters and firm distribution statistics suggest a much higher elasticity of trade flows to trade frictions than is estimated when trade flows are regressed on some series that is deemed to represent trade frictions. Further research is required on resolving this discrepancy, and this research is underway.

References

- ALESINA, A., E. SPOLAORE, AND R. WACZIARG (2005): “Trade, Growth and the Size of Countries,” chap. in: P. Aghion and S. Durlauf (ed.) “Handbook of Economic Growth”.
- ANDERSON, J., AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzle,” *The American Economic Review*.
- (2004): “Trade costs,” *Journal of Economic Literature*.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): “New Trade Models, Same Old Gains?,” *The American Economic Review*.
- BERNARD, A., J. EATON, J. JENSEN, AND S. KORTUM (2003): “Plants and Productivity in International Trade,” *The American Economic Review*.
- BERNARD, A., S. REDDING, AND P. SCHOTT (2007): “Comparative Advantage and Heterogeneous Firms,” *Review of Economic Studies*.
- DAVIS, D., AND J. HARRIGAN (2011): “Good jobs, bad jobs, and trade liberalization,” *Journal of International Economics*.
- DEMIDOVA, S. (2008): “Productivity Improvements And Falling Trade Costs: Boon Or Bane?,” *International Economic Review*.
- EATON, J., AND S. KORTUM (2002): “Technology, geography, and trade,” *Econometrica*.
- FELBERMAYR, G., AND B. JUNG (2012): “Unilateral Trade Liberalization in the Melitz Model: A Note,” *Economics Bulletin*.
- FEYRER, J. (2009): “Trade and Income – Exploiting Time Series in Geography,” *NBER Working Paper*.
- FRANKEL, J., AND D. ROMER (1999): “Does trade cause growth?,” *The American Economic Review*.
- GHIRONI, F., AND M. MELITZ (2005): “International Trade and Macroeconomic Dynamics with Heterogeneous Firms,” *The Quarterly Journal of Economics*.
- KRUGMAN, P. (1980): “Scale economies, product differentiation, and the pattern of trade,” *The American Economic Review*.
- MELITZ, M. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*.

A Appendix 1: Model Solution

Transformations

- We assume that the rest of the world is exogenous with respect to h and j . Therefore the equations which relate the variables in the rest of the world to themselves can be eliminated and replaced with the exogenous parameters Y_R, M_R, W_R which appear in the other the equations for h and j . The parameter S_R and the variable D_R drop out completely. We therefore have reduced the system to 10 equations in 10 endogenous variables with 14 parameters.
- We write the system in real terms, so that $x_i = X_i/P_i$, for all variables and parameters, $X \in \{Y, W, D\}$ & $i \in \{h, j, R\}$. The price indices are replaced with the relative price indices, $Q_{hj} = P_h/P_j$ and $Q_{hR} = P_h/P_R$
- It is found that we can redefine the endogenous variables by mutiplying through by a combination of parameters of the model without changing any of the interesting ratios given by the model. Many of the parameters can then be cancelled. This vastly reduces the dimensionality of the problem of calibrating the parameters of the model. The substitutions made are:

$$\begin{aligned}
 B &= \frac{\tilde{c}(1-\beta)}{b^k} \\
 x_i &= \tilde{x}_i c^{\frac{\mu-\theta}{\theta(1-\mu)}} B^{\frac{1}{1-\mu}} S_h^{-\frac{\mu}{1-\mu}}, \quad \forall x \in \{d, y, w\} \text{ and } \forall i \in \{h, j, R\} \\
 M_i &= \tilde{M}_i c^{-\frac{\mu-\theta}{\theta(1-\mu)}} B^{-\frac{1}{1-\mu}} S_h^{\frac{1}{1-\mu}}, \quad \forall i \in \{h, j, R\} \\
 Q_{hR} &= \tilde{Q}_{hR} \left(\frac{\tilde{M}_R \tilde{w}_R}{\tilde{y}_R} \right)^{-\frac{1}{2\mu-1}} \\
 s_j &= \frac{S_j}{S_h} \\
 \Delta_h &= \delta_{hR}^{1-\mu} \tilde{y}_R \left(\frac{\tilde{M}_R \tilde{w}_R}{\tilde{y}_R} \right)^{\frac{\mu}{2\mu-1}} \\
 \Delta_j &= \delta_{jR}^{1-\mu} \tilde{y}_R \left(\frac{\tilde{M}_R \tilde{w}_R}{\tilde{y}_R} \right)^{\frac{\mu}{2\mu-1}} \\
 \delta &= \delta_{hj} = \delta_{jh}
 \end{aligned}$$

Notice that we do not transform Q_{hj} . The variables that will be used in the model solution are the *tilde* versions, which are implicitly defined above. This means that absolute values for the model's endogenous variables cannot be seen from our calibration. What we can see are changes between two calibrations. Moreover all our targets are relative variables and so

we are not losing any information by making these change of variables.

Simulated Method of Moments

Given a vector of data targets, T_D , and their modelled equivalents, T_M . We adjust the parameters of the model until $T_M = T_D$, using a version of the Simulated Method of Moments:

- Choose the same number of data targets as unknown parameters, and observe T_D .
- Guess some initial parameters and solve for the model equilibrium, and so determine T_M .
- Let $Y \equiv T_D - T_M$.
- Perturb each of the unknown parameters around their current values and record the effect these perturbation have on the T_M vector in the square matrix, X (each column records $T_M(new) - T_M(old)$).
- The vector, N , of the number of times these parameter perturbations have to be applied in order to move the modelled targets to the data, satisfies $Y - XN = 0$ i.e. $N = X^{-1}Y$
- N implies some new parameter set which, if the equilibria are linear in parameter space, will give us $T_M = T_D$.
- Given non-linearity, changing the parameters by the amount implied by the calculated value of N will not actually take us to the point where $T_M = T_D$. However, this process is applied repeatedly until we converge upon the solution.

B Appendix 2: Details of the data

Aggregate Data

All data is for the year 2005

- Bilateral trade between Spain and Portugal in goods is acquired from the OECD "STAN Bilateral Trade Database".
- Bilateral trade between Spain and Portugal in services is acquired from OECD "Trade in Services by Partner Country". Spain does not report exports in services to Portugal in 2005. Exports from Spain to Portugal in services are acquired from Portugal's reported imports from Spain.
- Data for trade in goods and services is the sum of these two values. Everything is reported in dollars.
- Bilateral trade flows in goods and services between Catalonia and the rest of Spain/rest of world is acquired from the Statistical Institute of Catalonia (IDESCAT) 2005 input-output table.
 - Value of Catalan exports to rest of Spain is 35.52% of Catalan GDP
 - Value of Catalan exports to rest of the world is 30.4% of Catalan GDP
 - Value of Catalan imports from rest of Spain is 25.16% of Catalan GDP
 - Value of Catalan imports from rest of world is 36.06% of Catalan GDP
- GDP of Spain and Portugal is from the world bank (world development indicators).
- Eurostat reports that Catalonia's GDP is 18.7% of Spain's GDP

Data Summary (\$)		
	RoSpain (h) & Catalonia (j)	Spain(h) & Portugal (j)
Y_h	919,275,613,609	1,130,798,885,738
Y_j	211,523,272,129	191,847,858,529
X_j^h	53,223,709,552	20,973,098,000
X_h^j	75,144,306,877	12,245,288,721
X_R^h	223,308,603,857	266,645,329,000
X_R^j	64,309,823,143	41,046,170,331
X_h^R	280,503,308,146	344,529,506,279
X_j^R	76,271,486,854	50,534,969,000

Micro Data

- We use 2005 data from the Encuesta Sobre Estrategias Empresariales (ESSE) which surveys a representative sample of manufacturing firms in Spain with more than 10 employees. 1911 firms provide information about their sales.
- We define a firm to be Catalan if more than 50% of a firm's employment is based in Catalan plants. A firm is Rest of Spain if less than 50% of its employment is based in Catalonia plants. According to this criterion, 414 firms in our sample are Catalan and 1497 are Rest of Spain.
- The key statistics we are interested in are the percentage of Catalan and Rest of Spain firms who sell in the Rest of Spain and Catalonia respectively and the percentage of Catalan and Rest of Spain firms who export to the rest of the world.
- To find this, we use the variable "geographic range of the market." This variable tells us the 1st, 2nd, 3rd, 4th and 5th most important markets to the responding firm.
- The six potential responses are 1. Local, 2. Provincial, 3. Regional, 4. National, 5. Abroad, 6. Domestic and Abroad.
- We assume that a Catalan firm only sells in Catalonia ($\Phi_j^j < \phi < \Phi_h^j$) if the firm does not list "National," "Domestic and Abroad" and "Abroad" as one of the most important markets and if they say they do not export. We assume a Rest of Spain firm does not sell in Catalonia if the same conditions are met.
- A firm sells in the ROW if it responds yes to the question of whether or not they export.
- For verification of the calibration, we look at the ratio of sales of firms who at least sell in the rest of Spain $\phi > \Phi_h^j$ to sales of firms who export $\phi > \Phi_R^j$

Average Sales of Firms		
	Sales Revenue (€)	n
Firms with $\phi > \Phi_h^j$	95,404,810	355
Firms with $\phi > \Phi_R^j$	108,394,800	307
Ratio = 0.88		

- Domestic sales = total sales - export revenue. The standard deviation of log domestic sales is 1.9