THE OXFORD HANDBOOK OF

FINANCIAL REGULATION

Edited by

NIAMH MOLONEY, EILÍS FERRAN,

and

JENNIFER PAYNE

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2015

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2015934487

ISBN 978-0-19-968720-6

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

CHAPTER 18

CONDUCT OF BUSINESS REGULATION

ANDREW F TUCH*

I. Introduction 538
II. Regulatory Backdrop 539
1. Coexisting general law obligations 539
2. Economic and other justifications 541
3. Modal regulatory strategies 543
4. Complex regulatory frameworks 544
III. The Distinctive US Experience 547
1. A bifurcated regulatory regime 547
2. Divergent rules of conduct 549
3. Similar disclosure practices 552
4. Remuneration-based risks 553
5. Regulatory oversight 555
IV. International Comparisons 555
1. Standards of conduct 556
2. Remuneration-based risks 560
3. Enforcement and effectiveness 561
V. Financial Crisis and Other Recent Developments 561
VI. Conclusion 564

* For helpful comments on earlier drafts of this Chapter, I thank Deborah DeMott, Eilís Ferran, Howell Jackson, Arthur Laby, Don Langevoort, Niamh Moloney, Jennifer Payne, and Hillary Sale.

I. Introduction

Conduct of business (COB) regulation governs financial intermediaries' conduct toward their clients; that is, toward the actors—whether individuals or institutions—with whom financial intermediaries transact in providing financial products and services.[1] While the expression 'conduct of business regulation' is not widely employed in some jurisdictions, including the US, it is commonly used by international financial regulatory bodies and by financial regulators in many jurisdictions, including the Member States of the EU.[2] COB regulation governs financial intermediaries acting for or on behalf of their clients, such as in giving advice, exercising discretion, and executing orders. It may also govern intermediaries' arm's-length arrangements with clients—transactions in which intermediaries act as principals, or counterparties, in buying or selling financial products. COB regulation typically applies across the spectrum of financial intermediaries' functional lines of business, including their securities, banking, and insurance activities. It takes various forms, including requirements for registration or licensing; rules governing sales, marketing, and other business practices; and mechanisms of enforcement.[3]

COB regulation serves the objectives of protecting clients (investors) from harm, preserving and enhancing the integrity and orderly operation of financial markets, and otherwise serving the public interest. Because its focus is 'client-facing', it does not encompass 'firm-facing' regulation, such as the imposition of general supervision obligations, record-keeping requirements, or net capital requirements—regulation that nevertheless serves to protect clients.[4] Moreover, because of its focus on conduct, COB regulation does not encompass product regulation except to the extent such rules shape financial intermediaries' conduct toward their clients.

This Chapter considers various functional lines of business, but focuses on securities. As Professor Eddy Wymeersch observed, COB regulation is more symptomatic of securities regulation than of banking and insurance regulation.[5] The latter

[1] In this Chapter, references to 'financial intermediary' encompass both the firm itself and the individuals acting for it. The term 'client' is used to encompass all actors with whom financial intermediaries transact, whether they do so as principal or not, in providing financial products and services.

[2] Sometimes also referred to as 'business conduct' or 'market conduct' regulation, COB regulation is typically contrasted with market stability regulation and safety and soundness regulation. For prominent use of business conduct regulation in the US, see Department of the Treasury, Blueprint for a Modernized Financial Regulatory Structure (2008), 2–5, 14, 19–21, 138, 170–80.

[3] ibid, 171.

[4] The terms 'client-facing' and 'firm-facing' used in this Chapter are adopted from Moloney, N, *How to Protect Investors: Lessons from the EC and the UK* (2010).

[5] Wymeersch, E, 'The Structure of Financial Supervision in Europe: About Single Financial Supervisors, Twin Peaks and Multiple Financial Supervisors' (2007) 8 *European Business Organization Law Review* 237.

areas of financial regulation have been primarily concerned with the solvency of banks and insurance companies, whereas securities regulation has focused on investor protection—an objective served by COB regulation.[6]

The Chapter begins with the regulatory backdrop to COB regulation. It describes the justifications for COB regulation, the modal regulatory strategies used, and the complex frameworks within which COB regulation operates. The Chapter then generally assesses US COB regulation, outlining important market and regulatory developments over the past several decades, and drawing comparisons with corresponding EU and other COB regulation. The Chapter concludes by discussing reforms proposed or adopted in the wake of the global financial crisis of 2007–09.[7]

II. Regulatory Backdrop

1. Coexisting general law obligations

Financial regulation, and COB regulation in particular, is typically considered distinct from the general law, or private law.[8] The sources of financial regulation are legal instruments such as statutes and the rules and regulations of agencies, as well as judicial and other adjudicative opinions interpreting and applying these instruments. In the two jurisdictions with the deepest capital markets[9] and most important financial centres, namely the UK and the US, both COB regulation and the general law apply to regulate the business conduct of financial intermediaries.[10]

[6] See generally Jackson, H, 'Regulation in a Multisectored Financial Services Industry: An Exploratory Essay' (1999) 77 *Washington University Law Quarterly* 319, 348–52.

[7] In view of the complexity of the regulatory regimes considered, this Chapter should be regarded as an introduction to the field.

[8] The term 'general law' is used here to describe laws of general application, including contract law, property law, equity, and tort law. For similar distinctions, see Nelson, P, *Capital Markets Law and Compliance: The Implications of MiFID* (2008) 146, 306 (describing 'general law' in England and Wales); Hudson, A, *The Law of Finance* (2nd edn, 2013) 53–91 (contrasting EU and UK financial regulation with the 'substantive law' or 'private law' of England and Wales); DeMott, D and Laby, A, 'The United States of America' in Busch, D and DeMott, D, *Liability of Asset Managers* (2012) 411, 435–40 (discussing the 'private law' duties of investment advisers); and Baxt, R, Black, A, and Hanrahan, P, *Securities and Financial Services Law* (7th edn, 2008) 529–93 (contrasting Australian financial regulation with 'general law'). The boundaries of general law and financial regulation often blur. Hudson, n 8 above, 66–70.

[9] Cox, J et al., *Securities Regulation: Cases and Materials* (7th edn, 2013) 104.

[10] As to the UK, see Hudson, n 8 above, 53–66. As to the US, see Sitkoff, R, 'The Fiduciary Obligations of Financial Advisers under the Law of Agency' (2014) *Journal of Financial Planning* 42, 42 and SEC, Study on Investment Advisers and Broker-Dealers: As Required by Section 913 of the Dodd-Frank Wall Street Reform and Consumer Protection Act (2011), 45, 51.

The general law serves as an important backdrop against which to consider COB regulatory developments. Under general law, a financial intermediary providing financial products and services may face liability for fraud and for carelessness and disloyalty. Liability for fraud arises where the tort of deceit is committed.[11] Liability for carelessness arises from breach of a duty of care imposed by contract or the tort of negligence.[12] Liability for disloyalty arises from breach of a duty of loyalty, a duty arising where the financial intermediary–client relationship is characterized as fiduciary.[13]

Fiduciary doctrine emanates from the general law. The standard of propriety it imposes is unequalled elsewhere in the general law.[14] Generally speaking, fiduciary duties arise where one party has power or influence over the interests of another who is therefore vulnerable to the former party's exercise of discretion.[15] Some relationships in the financial context are, based on their status, fiduciary relationships. For instance, where a financial intermediary acts as the trustee of a pension fund, a scheme manager, or trustee of a unit trust, it will have fiduciary status.[16] Fiduciary duties may also arise on an ad hoc basis, such as where a financial intermediary provides advice or has discretionary control over a client's assets.[17] Where fiduciary duties arise, they demand 'undivided' or 'single-minded' loyalty, thus limiting the financial intermediary's pursuit of self-interest. Fiduciary duties do so by restraining the intermediary's freedom to act in ways inconsistent, or in conflict, with the interests of its client. In particular, fiduciary doctrine has traditionally required fiduciaries to avoid conflicts of interest, absent the informed consent of the party to whom the duty is owed.[18] In many jurisdictions, however, contracting parties may exclude or disclaim the existence of fiduciary duties.[19]

In the context of financial products and services, the general law alone is inadequate for regulating COB. The general law typically governs conduct in a less

[11] eg, Hudson, n 8 above, 714–30.

[12] Nelson, n 8 above, 306. In some common law jurisdictions, a duty of care may also arise from the fiduciary characterization of a relationship.

[13] *Bristol and West Building Society v Mothew* [1998] ch 1, 18.

[14] ibid, 16–19.

[15] Regarding unifying features of fiduciary relationships, see DeMott, D, 'Beyond Metaphor: An Analysis of Fiduciary Obligation' (1988) 37 *Duke Law Journal* 879, 902.

[16] Hudson, n 8 above, 104.

[17] eg, *Woods v Martins Bank Ltd* [1959] 1 QB 55; *Australian Securities and Investments Commission* (hereinafter, ASIC) *v Citigroup Global Markets Australia Pty Ltd* (No 4) (2007) 160 FCR 35.

[18] As to the position in England and Wales and in Australia, see *Bristol*, n 13 above, 18 and *Breen v Williams* (1996) 186 CLR 71, 113, 137–8. Fiduciary doctrine in the US varies by context, although the duty of loyalty on agents generally conforms to the Anglo-Australian approach of prohibiting conflicts of interest, absent informed consent. Restatement (Third) of Agency, section 8.01; and Sitkoff, n 10 above, 44.

[19] Contractual exclusion would seem permissible in the UK and Australia. Hudson, n 8 above, 112–15; *ASIC v Citigroup Global Markets Australia Pty Ltd* (No 4) (2007) 160 FCR 35. The position is more restrictive in the US. See Restatement (Third) of Agency, section 8.06. The formulation in section 8.06 accords with that recently employed by the Delaware Court of Chancery. See *In re Rural Metro Corp.*, 88 A3d 54, 101 (2014).

granular fashion than does regulation, which may impose specific, prescriptive obligations—an advantage where particular regulatory resolutions are considered desirable. The general law also adapts slowly and unpredictably to changes in market structure and to emerging market practices and technological developments, due to the nature of the judicial process, the generality of the legal concepts involved, and the limitations of judicial expertise. The general law provides no mechanism for public enforcement. It nevertheless applies broadly, thus filling some gaps left by COB regulation.

2. Economic and other justifications

The strategies used in COB regulation find theoretical justification in economics and related disciplines.[20] Standard neoclassical economic analysis asserts the need for regulatory intervention when particular market failures exist.[21] For instance, financial intermediaries with market power, perhaps resulting from high search costs,[22] may set higher prices than competitors. These circumstances would justify mandatory disclosures and direct price regulation.[23] Failures stemming from the 'public good' nature of information may be addressed by disclosure requirements and anti-fraud rules.[24] When a client delegates discretion to a financial intermediary to act on its behalf, (economic) principal-agent theory counsels for mechanisms to help align the intermediary's interests with those of the client and thereby to reduce 'agency costs'.[25] That theoretical framework justifies rules requiring loyalty, such as fiduciary duties.[26] The possibility that a financial intermediary will

[20] For an overview of justifications in the context of retail investors, see Moloney, n 4 above, 45–92. Policymakers often justify regulation on grounds of fairness, eg, Mary Jo White (Chair, SEC), Enhancing our Equity Market Structure, speech at Sandler O'Neil & Partners, LP Global Exchange and Brokerage Conference (5 June 2014), available at <http://www.sec.gov/News/Speech/Detail/Speech/1370542004312#.U6SU8qNLOAI> (last accessed 23 June 2014).

[21] Campbell, J et al., 'Consumer Financial Protection' (2011) 25 *Journal of Economic Perspectives* 91, 92–5.

[22] ibid, 92–3. [23] ibid.

[24] Campbell, J et al., The Regulation of Consumer Financial Products: An Introductory Essay with Four Case Studies (2010), Harvard Kennedy School Faculty Research Working Paper Series No 8, available at <http://dash.harvard.edu/bitstream/handle/1/4450128/Madrian_TheRegulationof.pdf?sequence=1>.

[25] Agency costs arise from the agent's divergence of interests from those of the principal. Jensen, M and Meckling, W, 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure' (1976) 3 *Journal of Financial Economics* 305, 308–10, 357.

[26] For a more extensive discussion of how (economic) principal-agent theory justifies the imposition of conflict of interest rules, such as duties of loyalty, see Cooter, R and Freedman, B, 'The Fiduciary Relationship: Its Economic Character and Legal Consequences' (1991) 66 *New York University Law Review* 1045; Sitkoff, R, 'The Economic Structure of Fiduciary Law' (2011) 91 *Boston University Law Review* 1041; Tuch, A, 'Conflicted Gatekeepers: The Volcker Rule and Goldman Sachs' (2012) 7 *Virginia Law and Business Review* 365, 378–92.

harm another through its carelessness justifies the imposition of a duty of care.[27] These various justifications have greater force with regard to regulation to protect individual (or retail) investors than that to protect institutional investors, because institutional investors have been considered sophisticated and thus able to 'fend for themselves', such as by contracting with financial intermediaries to protect their financial interests. However, the notion of institutional investors as adequately able to guard their own interests stands on weaker foundations after the financial crisis, when abundant evidence inconsistent with this notion surfaced.[28]

Some strategies used in COB regulation may also be justified by evidence indicating that clients, particularly retail clients, fail to conform to the behavioural assumptions made in neoclassical economics. Investors have biases and cognitive limitations, leading them to make decisions that systematically depart from rationality.[29] Cognitive psychology has demonstrated the tendency of individuals to be overconfident in their judgements, abilities, and prospects; to hold onto opinions 'too tightly and for too long'; to be more disposed towards avoiding a loss than taking a gain; and to anchoring any estimates they form to some initial, possibly arbitrary value.[30] Applying these insights to investor behaviour, behavioural finance has demonstrated seemingly irrational behaviour by investors. A flourishing legal literature has drawn upon insights about the behaviour of investors and cautiously suggested new directions for regulatory intervention.[31]

Importantly, even where regulatory intervention finds theoretical justification, determining the best form of such regulation poses difficult challenges. The fit between economic and other justifications and any resulting regulation is, at best, imprecise.[32] Moreover, the task of crafting desirable COB regulation occurs against the backdrop of the general law. Richard Posner has observed, in general, that any 'market failures' are those of the market and of the rules prescribed by the general law.[33] In

[27] See Shavell, S, *Foundations of Economic Analysis of Law* (2004) 178–81 (referring to firms generally rather than financial intermediaries specifically).

[28] See Langevoort, D, 'Global Securities Regulation after the Financial Crisis' (2010) *Journal of International Economic Law* 799, 809–11; and Langevoort, D, 'The SEC, Retail Investors, and the Institutionalization of the Securities Markets' (2009) 95 *Virginia Law Review* 1025, 1058, 1061–70.

[29] For an overview of the literature in behavioural economics, see Barberis, N and Thaler, R, 'A Survey of Behavioral Finance' in Constantinides, G et al. (eds), *Handbook of the Economics of Finance* (2003) 1054.

[30] ibid, 1065–9. See also Campbell, n 21 above, 94.

[31] Langevoort, D, 'Taming the Animal Spirits of the Stock Markets: A Behavioral Approach to Securities Regulation' (2002) 97 *Northwestern University Law Review* 135; Langevoort, D, 'Selling Hope, Selling Risk: Some Lessons for Law from Behavioral Economics about Stockbrokers and Sophisticated Customers' (1996) 84 *California Law Review* 627; Prentice, R, 'Whither Securities Regulation' (2002) 51 *Duke Law Journal* 1397, 1448–89.

[32] A further difficulty for crafting regulatory strategies stems from the potential behavioural biases of regulators. Choi, S and Pritchard, A, 'Behavioral Economics and the SEC' (2003) 56 *Stanford Law Review* 1, 20–41.

[33] Posner, R, *Economic Analysis of Law* (7th edn, 2007) 389 (referring to common law, rather than general law).

the present context, any failures are those of the market and of the rules prescribed by both the general law and existing COB regulation. Where such failures arise, regulatory reforms must be assessed for their likely effect on people and markets, an exercise often involving expert judgements and contested evidence.[34]

3. Modal regulatory strategies

In regulating COB, major regimes employ a range of regulatory strategies that broadly map onto the general law obligations, but with important differences. The primary strategies are anti-fraud rules and duties of care, loyalty, fair dealing, and best execution—as well as variants of these duties.[35] Other core regulatory strategies include registration or licensing requirements and mechanisms to enforce the duties imposed.

By way of general explanation, anti-fraud rules create a cause of action against parties intentionally engaging in misleading or deceptive conduct.[36] Those rules are often broader than general law fraud rules since they extend beyond intentional affirmative misrepresentations to encompass the failure to disclose material adverse facts.[37] Disclosure requirements promote information production and may differ along a variety of dimensions. Some require highly particularized information about products, services, or intermediaries. Others may be tailored to individual transactions or to the aggregate of transactions in which an intermediary is involved.[38] Duties of best execution concern the execution of trades for clients, including the handling of clients' orders.

Duties of care typically require that the process employed by a financial intermediary in giving advice or making a recommendation be 'suitable'. For instance, the US imposes suitability duties on broker-dealers and investment advisers, as does the EU on investment firms. The concept of 'suitability' draws on certain characteristics of the client involved and the securities or investment strategy under consideration. Most jurisdictions apply somewhat weaker suitability duties for the benefit of institutional clients. In the EU, regulatory strategies regarding suitability are explicitly tailored according to clients' categorization as retail investors, professional investors, or eligible counterparties. Unlike the general law, some jurisdictions also supplement their duties of care with detailed evidential requirements, for instance requiring firms to document their advice to clients.

[34] Breyer, S, *Regulation and its Reform* (1982) 184–8, 191–6.

[35] For a general discussion of the range of regulatory strategies available, see Campbell, n 24 above, 14–19.

[36] However, some anti-fraud provisions do not require scienter, eg, Investment Advisers Act of 1940, section 206(2); and Securities Act of 1933, section 17(a)(2) and (3).

[37] *Hanly v SEC*, 415 F.2d 589, 592 (2d Cir. 1969).

[38] Campbell, n 24 above, 15.

Rules requiring loyalty—that is, rules regulating conflicts of interest—are essential to COB regulation, due in particular to the remuneration-based risks many financial intermediaries face and the organizational structure they employ.[39] But rather than mandate conflict avoidance, COB regulation typically requires financial intermediaries to 'manage' conflicts of interest, especially by disclosing them, except in more egregious situations in which certain conflicts of interest require client consent or are banned. Given the difficulties involved in determining whether a financial intermediary exploits, or acts upon, any conflicting interests it faces, regimes typically leave the task of 'managing' conflicts to firms themselves and provide limited guidance on the meaning of conflict management. Conflict of interest rules in COB regulation thus differ significantly from general law fiduciary duties—but they may not be weaker, since they typically cannot be contractually disclaimed.[40]

Other modal regulatory strategies are registration or licensing requirements and mechanisms of public enforcement. Most COB regimes require financial intermediaries to register with regulatory bodies before they may engage in financial activities with clients, unless an exemption applies. To register, an intermediary must pass certain financial capital and competence tests. Public enforcement ensures that widespread practices may be tackled on a systematic basis by a regulator with access to an array of deterrent mechanisms, among the most powerful of which is the ability to suspend or revoke an intermediary's registration privileges. These modal strategies have no general law counterparts.

4. Complex regulatory frameworks

Many of the frameworks in which COB regulation operates are complex. The US approach involves multiple layers of rules, sources of law, and regulators. The product is a complex and often esoteric amalgam of laws.[41] For example, determining COB regulation for broker-dealers requires resort to federal and state statutes, the rules and regulations of federal and state public regulators, rules and interpretations of self-regulators, and formal and informal pronouncements of regulators.[42] Such regulation coexists with federal and state general law. Enforcement is undertaken by federal and state regulators, by self-regulators, and even by federal and state criminal prosecutors. Multiple private actions may also be under way. The framework may thus give rise to simultaneous, uncoordinated proceedings.[43]

[39] See Sections III.4 and IV.2. [40] eg, Securities Exchange Act of 1934, section 29.

[41] Langevoort, D, 'Brokers as Fiduciaries' (2010) 71 *University of Pittsburgh Law Review* 439, 443 ('Why is this area of the law [concerning broker-dealers] so confusing?').

[42] See Section III. See also DeMott and Laby, n 8 above, 412.

[43] One instance of cooperation among regulators was the 2002 $1.4 billion 'Global Settlement' among various agencies with major financial conglomerates regarding research analysts' skewed

Widely diverging approaches are adopted across securities, futures, insurance, and banking sectors—with different regulators often dedicated to particular pieces of the financial regulatory puzzle.[44]

On top of this, the US regulatory approach reflects piecemeal, incremental reform, rather than coherent, wholesale reinvention. Despite seismic changes introduced in the wake of the global financial crisis by the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010[45] (hereinafter, the Dodd–Frank Act), the US Depression-era regulatory framework remains largely intact. Reforms have typically responded to high-profile issues of the day. In the securities industry, these have included the receipt of 'soft dollar' benefits by investment advisers, the incorporation of retail brokerage into financial conglomerates, the rise of proprietary trading, and research analyst conflicts of interest.[46] Many such issues reflect competitive pressures that have developed following the end of fixed brokerage commissions in 1975.[47] This piecemeal approach seems likely to continue as regulators grapple with the proliferation of trading venues for securities and the complex order types that these venues offer, as well as the increasing use of algorithmic trading strategies by high-frequency and other trading firms.

The regimes of Member States of the EU, such as the UK, are also complex, stemming in part from the rule harmonization process and the need to accommodate the different legal traditions, regulatory styles, and underlying market practices of the Member States. The EU relies on a complex, often overlapping and occasionally underlapping, patchwork of legal instruments. Numerous directives govern COB regulation.[48] For securities, or investments, the 2004 Markets in Financial Instruments Directive[49] (hereinafter, MiFID I) was implemented by the EU Member States in 2007 to regulate the provision of investment services and activities, including investment advice, portfolio management, and trade execution, in respect of a broad range of financial instruments.[50] While MiFID I will be repealed by the recently agreed 2014 MiFID II/MiFIR regime, which will begin applying in

research reports. SEC Press Release: SEC, NY Attorney General, NASD, NASAA, NYSE and State Regulators Announce Historic Agreement to Reform Investment Practices (20 December 2002).

44 Department of the Treasury, n 2 above, 25–61; Coffee, J and Sale, H, 'Redesigning the SEC: Does the Treasury Have a Better Idea?' (2009) 95 *Virginia Law Review* 707.

45 Pub. L. No. 111–203, 124 Stat. 1376 (2010).

46 See Langevoort, n 41 above, 440 (identifying several high-profile issues).

47 ibid.

48 As to its overlap with other EU financial services directives, see Linklaters, MiFID II: Key Interactions between MiFID/MiFIR II and Other EU and US Financial Services Legislation (2012) ('[t]hese overlaps result in some cases in conflicting obligations that are impossible to comply with …').

49 MiFID I includes Directive 2004/39/EC [2004] OJ L 145/1 (hereinafter, MiFID I or MiFID I Level 1) and detailed administrative rules under Directive 2006/73/EC [2006] OJ L 241/26 implementing Directive 2004/39/EC and Commission Regulation (EC) No 1287/2006 [2006] OJ L 241/1 (hereinafter, MiFID I Level 2).

50 MiFID I regulates investment firms, which are legal persons whose regular business is the provision of investment services to third parties and/or the performance of investment activities on a

2017, the main pillars of COB regulation will remain largely unchanged.[51] The distribution of some insurance-based investment products that fall outside MiFID I is regulated by the Insurance Mediation Directive,[52] and is thus subject to somewhat lighter COB regulation, despite these products' functional equivalence to products falling under MiFID. MiFID I also focuses on duties, and leaves questions of liability for breaches of those duties to be determined at the national level.

Complexities also exist at the national level. The UK implemented MiFID I, as required under EU law, but has relied on Article 3 to 'opt out' with regard to some of its smaller firms.[53] Accordingly, the UK's current COB Sourcebook prescribes rules for both MiFID and non-MiFID firms and business. For non-MiFID firms and business, the UK regulator has had to determine whether to apply MiFID I requirements and definitions. Doing so would harmonize regulation and minimize opportunities for regulatory arbitrage and yet would not be as suited to local conditions as are tailored rules.[54] The UK also has had to decide whether to 'gold-plate' any of the MiFID I requirements—the practice (discouraged by MiFID I) of Member States imposing additional or stricter obligations on local markets and actors beyond the MiFID I requirements.[55] Additionally, the UK must accommodate its 2013 Retail Distribution Review reforms, which banned the payment of certain commissions, into its EU obligations.[56] Despite these challenges, which are replicated to different extents across the Member States, there seems to be increasing consistency among EU Member States, with some Member States applying MiFID I principles to firms and financial instruments outside MiFID I's scope.[57] Of course, in the UK (as in all the Member States), COB regulation also coexists with the general law.[58]

professional basis. MiFID I defines investment services and activities to include investment firms' investment advice, as well as portfolio management and order execution they perform on behalf of clients. Other activities include the reception and transmission of orders, underwriting, and the operation of multilateral trading facilities. MiFID I applies to a broad range of financial instruments, but excludes deposit-based investments and unit-linked insurance investments. See MiFID I Level 1, Article 4 and Annex I.

[51] 2014 MiFID II includes Markets in Financial Instruments Directive 2014/65/EU (hereinafter, 2014 MiFID II) ([2014] OJ L 173/349) and MiFIR includes Markets in Financial Instruments Regulation (EU) No 600/2014 (MiFIR) ([2014] OJ L173/84).

[52] Directive 2002/92/EC [2003] OJ L9/3.

[53] MiFID I includes an optional exemption in Article 3 for firms that do not hold client assets or funds and only advise on and transmit orders for certain financial instruments. This option will be retained under the MiFID II/MiFIR regime, although its availability has been narrowed.

[54] Financial Services Authority, Reforming Conduct of Business Regulation (2006), Consultation Paper 06/19, 12–16.

[55] MiFID I Level 2, Article 4. Conditions apply to Member States that seek to retain or impose COB rules additional to those governed by the Article 4 'gold-plating' ban.

[56] Financial Services Authority, A Review of Retail Distribution (2007), Discussion Paper 07/1.

[57] The UK has taken advantage of Article 3 and imposes its own regulatory regime on Article 3 firms. This regime is 'closely based on MiFID requirements'. Moloney, n 4 above, 24.

[58] See n 8 above.

III. The Distinctive US Experience

This Section describes the distinctive US experience with COB regulation in the field of securities. That experience involves a bifurcated structure adopted in the aftermath of the market collapse of 1929 and Great Depression.[59]

1. A bifurcated regulatory regime

US federal securities law requires financial intermediaries in the business of providing securities-related services,[60] including advice and recommendations, to register with the Securities and Exchange Commission (SEC), unless they are exempt from registration or otherwise not required to register.[61] There are two broad categories of registrant: investment advisers and broker-dealers. Investment advisers are those who, for compensation, are in the business of providing advice, or issuing reports or analyses, regarding securities.[62] They give advice 'for its own sake' and are compensated specifically for that advice,[63] and, importantly, are generally remunerated on the basis of funds under management.[64] They are regulated by the SEC pursuant to the Investment Advisers Act of 1940 (hereinafter, the Advisers Act).[65]

Investment advisers advise both retail and institutional clients. Among their institutional clients are collective investment schemes, or pooled investment vehicles, such as mutual funds, private equity funds, and hedge funds. Where these funds are offered to the public, they are typically regulated by the Investment Company Act of 1940. The advisers to these funds will, nevertheless, be investment advisers and thus subject to the Advisers Act.

In contrast, broker-dealers are those acting as brokers (in the business of 'effecting transactions in securities' for others[66]) or as dealers (in the business of 'buying

[59] For a discussion of market and other conditions leading to reforms, see Loss, L and Seligman, J, *Securities Regulation* (3rd edn, Vol 1, 1998) 166–272.

[60] The term 'securities' is defined in section 2(a)(1) of the Securities Act of 1933 to include notes, stocks, bonds, debentures, investment contracts, and 'any interest or instrument commonly known as a "security"'. Although the definition has produced an unsettled body of law, Cox, n 9 above, 27–8, it is broad, capturing interests in mutual funds, hedge funds, variable insurance products, and exchange traded funds, among other investment vehicles. SEC, n 10 above, 65–6.

[61] In the US, the Commodity Futures Trading Commission has regulatory authority over on-exchange traded futures and over-the-counter (OTC) derivatives (swaps). The relevant self-regulatory organization is the National Futures Association.

[62] Investment Advisers Act of 1940, section 202(a)(11).

[63] *Thomas v Metropolitan Life Insurance Co.*, 631 F.3d 1153, 1166 (2011).

[64] SEC, n 10 above, iii.

[65] The SEC focuses its regulatory attention on investment advisers managing more than $25 million or associated with a mutual fund. That limit was increased to $100 million by the Dodd–Frank Act.

[66] See n 40 above, section 3(a)(4)(A).

and selling securities' on their own behalf[67])—or as both brokers and dealers. Regulated by the Securities Exchange Act of 1934 (hereinafter, the Exchange Act), broker-dealers may give advice or make recommendations about securities (and have done so increasingly in recent decades[68]), but they primarily perform other securities-related functions, including executing client trades and providing generalized or client-specific research. They typically receive transaction-based compensation, such as commissions.[69] Broker-dealers that give advice fall outside the definition of investment adviser provided their investment advice is 'solely incidental to the conduct of [their] business as a broker or dealer' and they receive 'no special compensation' for that advice.[70] Those broker-dealers that must register as investment advisers are dual-registered. Only about 5 per cent of investment advisers are dual-registered, but that number includes nearly all of the largest retail broker-dealers.[71]

Investment advisers and broker-dealers are subject to distinct regulatory regimes. Under the Advisers Act and Exchange Act, respectively, it is generally unlawful for a person to act as an investment adviser or a broker-dealer without being registered with the SEC.[72] In addition, broker-dealers that deal with the public must register with the Financial Industry Regulatory Authority (FINRA),[73] the self-regulatory organization that in 2007 succeeded to the functions of the National Association of Securities Dealers and the regulatory arm of the New York Stock Exchange (NYSE). No industry regulator exists for investment advisers, and separate divisions of the SEC regulate each type of registrant. Accordingly, broker-dealers are subject to both SEC and FINRA regulation and enforcement,[74] while investment advisers face only SEC regulation and enforcement. State registration requirements also apply to both.[75]

Although they are subject to distinct regimes, investment advisers and broker-dealers face broadly similar regulatory strategies in most respects. Both are

[67] ibid, section 3(a)(5)(A).

[68] Laby, A, 'Reforming the Regulation of Broker-Dealers and Investment Advisers' (2010) 65 *The Business Lawyer* 395, 398.

[69] SEC, n 10 above, 7.

[70] 15 U.S.C. section 80b-2(a)(11)(C).

[71] SEC, Staff Study on Enhancing Investment Adviser Examinations as Required by Section 914 of the Dodd–Frank Wall Street Reform and Consumer Protection Act (2011), 37.

[72] Investment Advisers Act of 1940, section 203; See n 40 above, section 15(a).

[73] SEC, n 10 above, 47. Broker-dealers may also choose to become members of a national securities exchange, such as the NYSE (the rules of which FINRA enforces).

[74] FINRA is regarded as 'the best first line defense against unethical or illegal securities practices' by broker-dealers. *First Jersey Securities, Inc. v Bergen*, 605 F.2d 690, 698–9 (3d Cir. 1979). FINRA writes rules and enforces them, and rivals the SEC in terms of its budget and personnel. See Irwin, S et al., 'Self-Regulation of the American Retail Securities Markets—an Oxymoron for What Is Best for Investors?' (2012) 14 *University of Pennsylvania Journal of Business Law* 1055, 1073.

[75] But federal law is the primary concern. In 1996, Congress passed the National Securities Markets Improvement Act of 1996 to exempt broker-dealers and investment advisers from states' registration procedures in important contexts. See also n 65 above.

subject to anti-fraud rules.[76] Both owe duties of best execution as well as duties of loyalty, of good faith, and to use particular standards of care.[77] Nevertheless, the rules applicable to each type of registrant differ, although both are subject to section 10(b) and related Rule 10b-5 of the Exchange Act, probably the most formidable anti-fraud provisions in the SEC's regulatory arsenal.[78]

2. Divergent rules of conduct

While the primary difference between the US COB obligations of investment advisers and broker-dealers is often said to be the fiduciary status of the former, rule differences run deeper and are best considered separately with regard to the duties of loyalty and care. Much of the scholarly research has focused on these differences.[79]

Regarding duties of loyalty, investment advisers have the status of fiduciaries, pursuant to the Supreme Court's interpretation in *SEC v Capital Gains Research Bureau, Inc.*[80] of section 206 of the Advisers Act, an anti-fraud provision. According to the Supreme Court, the Advisers Act 'reflects a congressional recognition of the delicate fiduciary nature of an investment advisory relationship, as well as a congressional intent to eliminate, or at least to expose, all conflicts of interest which might incline an investment adviser—consciously or unconsciously—to render advice which was not disinterested'.[81] Broker-dealers, in contrast to investment advisers, do not enjoy the status of fiduciaries, but may be fiduciaries on an ad hoc basis, where the facts and circumstances justify it. Although the exercise of discretion over client assets may well justify fiduciary characterization,[82] judicial decisions are difficult to reconcile and no clear consensus exists as to when broker-dealers owe fiduciary duties.[83]

[76] eg, section 15(c) of the Exchange Act and section 206 of the Advisers Act.

[77] The duty of fair dealing may capture myriad other misconduct, including overcharging, engaging in high-pressure sales techniques, and 'churning'. Broker-dealers' duty of fair dealing derives from anti-fraud provisions of federal securities laws and the so-called shingle theory, under which broker-dealers, by holding themselves out to the public as broker-dealers, make an implied representation that they will deal fairly with their clients. eg, *Charles Hughes & Co. v SEC*, 139 F.2d 434 (2d Cir. 1943), *cert. denied*, 321 U.S. 786 (1944).

[78] Under Rule 10b-5, both investment advisers and broker-dealers are subject to rights of action (both public and private) for material misstatements and omissions made with scienter in connection with the purchase or sale of a security.

[79] eg, Langevoort, n 31 above; Laby, n 68 above; Prentice, n 31 above.

[80] *SEC v Capital Gains Research Bureau, Inc.*, 375 U.S. 180 (1963).

[81] ibid, 191 (internal quotations and citations omitted).

[82] Hazen, T, 'Are Existing Stock Broker Standards Sufficient? Principles, Rules, and Fiduciary Duties' (2010) *Columbia Business Law Review* 710, 737–49; and SEC, n 10 above, 54–5.

[83] Coffee, J and Sale, H, *Securities Regulation: Cases and Materials* (12th edn, 2012) 661; and Cox, n 9 above, 1031.

The duty of care required of broker-dealers is better articulated than that imposed on investment advisers. Broker-dealers are subject to a so-called 'suitability' duty: a duty initially developed some 70 years ago to 'neutralize' the incentives broker-dealers have, by virtue of their remuneration structure, to skew their advice to generate commissions.[84] Today, the main duty stems from FINRA Rule 2111—although a narrower duty also stems from the anti-fraud provisions of the Exchange Act.[85] FINRA also imposes heightened suitability rules for certain activities or products, including variable annuities, penny stocks, day trading, and complex or particularly risky securities.[86] Broker-dealers also owe a 'know-your-customer' duty in their initial dealings with a client,[87] and a duty of fair dealing.

In relevant part, FINRA Rule 2111 requires a broker-dealer to have a 'reasonable basis to believe' that its recommendation is 'suitable' for both some investors (based on having conducted a reasonable investigation) and the particular client involved (based on that client's investment profile).[88] The duty cannot be satisfied by simply disclosing the risk[89] or even by ensuring a client understood the recommendation and decided to follow it,[90] and it cannot be contractually disclaimed.[91] Instead, the issue is whether, based on the information available to the broker-dealer, the recommendation was 'suitable'. There is no requirement to document the process for each client, even though a broker-dealer has a general obligation to evidence compliance with the suitability duty.[92] A broker-dealer may satisfy its reasonable investigation obligation by relying on the client's responses, unless 'red flags' exist regarding the accuracy of the information given to the client or the client's understanding of that information.[93] Since 1996, broker-dealers have owed a suitability duty to their institutional clients, although the duty is more easily satisfied than the one owed to retail clients.[94]

[84] Wrona, J, 'The Best of Both Worlds: A Fact-based Analysis of the Legal Obligations of Investment Advisers and Broker-Dealers and a Framework for Enhanced Investor Protection' (2012) 68 *The Business Lawyer* 1, 20–1.

[85] Unlike the duty deriving from the anti-fraud provisions, the duty deriving from FINRA rules does not require proof of scienter. *In the Matter of the Application of Jack H Stein*, Exchange Act Release No 47335 (10 February 2003).

[86] SEC, n 10 above, 65–6.

[87] FINRA Manual, Rule 2090.

[88] FINRA Manual, Rule 2111. In addition to reasonable basis and customer-specific components (described above), there is also a quantitative component to the suitability duty, which may be violated where excessive trading activity is recommended.

[89] *In the Matter of the Application of Jack H Stein*, n 85 above.

[90] eg, *In the Matter of the Application of Clinton Hugh Holland*, Exchange Act Release No 36621 (21 December 1995), at 10, aff'd, 105 F.3d 665 (9th Cir. 1997).

[91] FINRA Manual, Rule 2111.02.

[92] According to FINRA, the 'extent to which a firm needs to document its suitability analysis depends on an assessment of the customer's investment profile and the complexity of the recommended security …'. See Notice 12-25 (2012) 9.

[93] FINRA Notice 12–25, May 2012, 11.

[94] A broker-dealer satisfies the customer-specific component of its suitability requirement where it has a reasonable basis to believe that the institutional client is capable of evaluating the investment

The 'know your customer' duty requires broker-dealers to document a wide range of customer information at the time of opening an account. Specifically, it requires broker-dealers to use 'reasonable diligence' in opening and maintaining every client account to know the 'essential facts' about their clients.[95]

Investment advisers also owe a duty of care stemming from their fiduciary status as well as a separate suitability duty.[96] The duties of care and suitability incorporate process-based standards of care. In particular, investment advisers' duty of care has been described as comprising an 'affirmative duty of utmost good faith, and full and fair disclosure of all material facts, as well as an affirmative obligation to employ reasonable care to avoid misleading their clients'.[97] The SEC has described the duty as requiring investment advisers to serve their clients' 'best interests'[98] and, more specifically, as requiring 'a reasonable investigation' to ensure recommendations are not based on materially inaccurate or incomplete information.[99] To satisfy the duty of suitability, investment advisers' advice must be the result of a reasonable determination, taking account of the client's financial situation and investment objectives.[100] Neither duty has been well developed by the courts or the SEC.[101]

Broker-dealers owe obligations that may straddle duties of care and loyalty. They must 'observe high standards of commercial honor and just and equitable principles of trade', under FINRA Rule 2010.[102] FINRA may invoke the rule to sanction any conduct, whether or not it is caught by a specific rule or amounts to fraud. That catch-all rule allows FINRA to regulate the 'ethics and morality' of broker-dealers; a role that FINRA's predecessor was regarded as more capable of performing than government.[103] Broker-dealers also owe an obligation of fair dealing. Derived from statutory anti-fraud provisions, the obligation arises from the 'implied representation' that, in hanging out its shingle, a broker-dealer will deal fairly with its customers.[104] Violations of the obligation are typically avoided through disclosure. Commentators

risks and the client affirmatively indicates that it is exercising independent judgement. See FINRA Manual, Rule 2111(b).

[95] Essential facts include those required to, among various objectives, effectively service the client's account and comply with relevant laws. FINRA Manual, Rule 2090.

[96] SEC, n 10 above, 27; Hazen, T, *Treatise on the Law of Securities Regulation* (2014), section 21.4 ('[T]he [SEC] has taken the position that the antifraud provisions of the Investment Advisers Act can be used to enforce a suitability requirement').

[97] *SEC v Capital Gains Research Bureau, Inc.*, 375 U.S. 180, 191–2 (1963) (internal quotations omitted).

[98] eg, Proxy Voting by Investment Advisers, Investment Advisers Act Release No 2106 (2003).

[99] Concept Release on the U.S. Proxy System, Investment Advisers Act Release No 3052 (2010), 119.

[100] SEC, n 10 above, 27–8.

[101] Wrona, n 84 above, 11, 13, and 50–2; and SEC, n 10 above, 123.

[102] FINRA Manual, Rule 2010.

[103] See Seligman, S, *The Transformation of Wall Street* (3rd edn, 2003) 185–6 (quoting William O Douglas, then Chairman of the SEC).

[104] eg, *Charles Hughes & Co. v SEC*, 139 F.2d 434, 436–7 (2d Cir. 1943), *cert. denied*, 321 U.S. 786 (1944).

question the continued viability of this obligation due to the tacit nature of the representation, but the widespread use by broker-dealers of mandatory pre-dispute arbitration clauses in customer agreements has prevented contemporary judicial reconsideration of the obligation.[105] While no explicit counterpart obligations exist for investment advisers, their fiduciary duty may well require equivalent standards of conduct.

3. Similar disclosure practices

Despite their divergent rules of conduct, investment advisers and broker-dealers adopt remarkably similar approaches to conflicts of interest. As explained, investment advisers must either eliminate or disclose material conflicts of interest[106]—and their business interests will usually dictate they take the latter approach. Broker-dealers must also disclose conflicts of interest, such as adverse facts relevant to any recommendations they make, a duty stemming from anti-fraud provisions of the federal securities laws.[107] This disclosure duty for broker-dealers applies even in the absence of any fiduciary relationship[108] and even where their recommendations are suitable.[109] Investment advisers and broker-dealers are also subject to identical requirements to 'establish, maintain and enforce' information barriers.[110]

According to the SEC, the distinction between the regimes for regulating conflicts of interest for broker-dealers and investment advisers boils down to a difference in disclosure practices. With some exceptions,[111] broker-dealers and investment advisers must disclose the conflicts of interest they face—a practice that allows investment advisers to discharge their fiduciary duties and broker-dealers to avoid violating anti-fraud provisions of the Exchange Act. Nevertheless, the extent, form,

[105] See Karmel, R, 'Is the Shingle Theory Dead?' (1995) 52 *Washington and Lee Law Review* 1271, 1284–97.

[106] SEC Release No IA-2333; File No. S7-30-04 Registration under the Advisers Act of Certain Hedge Fund Advisers, citing *SEC v Capital Gains Research Bureau, Inc.*, 375 U.S. 180, 191–4 (1963); SEC, n 10 above, iii.

[107] *In the Matter of Richmark Capital Corp.*, Exchange Act Release No 48758 (7 November 2003) (Commission opinion). Some specific rules also require disclosure of conflicts of interest, eg, Exchange Act Rules 10b-10.

[108] *Leib v Merrill Lynch Pierce Fenner & Smith*, 461 F.Supp. 951, 953 (E.D. Mich. 1978).

[109] SEC, n 10 above, 103.

[110] Investment Advisers Act, section 204A; and n 40 above, section 15(g).

[111] In certain instances, investment advisers are clearly subject to stricter requirements than broker-dealers. First, before trading as principal with a client, an investment adviser must disclose the conflict and obtain its client's consent on a trade-by-trade basis; in contrast, broker-dealers must only disclose the capacity in which they act in the transaction confirmation note. SEC, n 10 above, 119. Second, in recommending proprietary products, investment advisers must disclose the existence of their skewed incentives, whereas broker-dealers need not. *Thomas v Metropolitan Life Insurance Co.*, 631 F.3d 1153 (2011).

and timing of the required disclosures differ between broker-dealers and investment advisers,[112] with investment advisers tending to need to disclose conflicts of interest more often and in greater detail than broker-dealers.[113] Investment advisers largely satisfy their duties of disclosure at the outset of relationships, and annually thereafter, through the use of 'disclosure brochures' (on form ADV).[114] Broker-dealers' disclosures generally need not be written and are typically made during the course of client relationships and on confirmation of transactions.[115] Accordingly, the different fiduciary characterization of investment advisers results in different disclosure practices, not necessarily a stricter standard of loyalty. Given the limitations behavioural finance has shown of the effectiveness of disclosures in sanitizing conflicts of interest,[116] the differences may mean little to investors in practical terms.

4. Remuneration-based risks

Broker-dealers face acute remuneration-based risks. Commission-based remuneration poses a particularly severe risk to the quality of an intermediary's advice. It produces incentives for the intermediary to maximize its commissions and thereby potentially leads to conduct inconsistent with a client's best interests, such as the provision of skewed advice. Accordingly, even though investment advisers and broker-dealers are subject to similar disclosure obligations, broker-dealers (paid by commission) would seem more likely to engage in disloyal conduct—a prediction borne out by the many instances of fraud by rogue broker-dealers.[117] Little in broker-dealer regulation specifically combats these remuneration incentives—and no change is on the horizon.

A further remuneration-based risk concerns third-party payments, or kickbacks, to financial intermediaries advising clients regarding choices from among a range of possible products and services. Professor Howell Jackson refers to this phenomenon as the trilateral dilemma.[118] The dilemma arises because such

[112] ibid, 114. [113] SEC, n 10 above, 106.

[114] But such disclosures may not satisfy investment advisers' disclosure obligations in all cases. SEC, n 10 above, 18 and 23.

[115] ibid, 106. Self-regulatory rules also impose disclosure obligations in particular contexts, eg, FINRA Rule 5121 and NASD Rule 2711.

[116] For evidence suggesting the disclosure of conflicts of interest inadequately protects those to whom the disclosure is made and may even lead to increased bias, see Cain, D et al., 'The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest' (2005) 34 *Journal of Legal Studies* 1. However, institutional measures may render disclosure effective in dampening adviser bias. See Church, B and Kuang, X, 'Conflicts of Interest, Disclosure and (Costly) Sanctions: Experimental Evidence' (2009) 38 *Journal of Legal Studies* 505.

[117] Langevoort, n 31 above, 630 (asking why there are 'so many notorious examples of broker cheating').

[118] Jackson, H, 'The Trilateral Dilemma in Financial Regulation' in Lusardi, A (ed.), *Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs* (2008) 82.

payments from third parties may skew advisers' advice, producing suboptimal outcomes for clients. However, as Professor Jackson observes, these arrangements may also represent efficient market mechanisms for financing the cost of distributing products and services. The issue has arisen in securities, banking, and insurance contexts, and it includes investment advisers' receipt of 'soft dollar' benefits in return for paying above-market commissions to broker-dealers for executing trades as well as broker-dealers' receipt of payments for directing 'order flows' to securities markets. A vast scholarly literature has resulted.[119] The US regulatory approach has been piecemeal, adopting a broad array of regulatory tools to address the dilemma where it arises. However, the modal response has been the imposition of some sort of fiduciary duty together with disclosure to affected clients.[120]

Section 28(e) of the Exchange Act regulates the receipt by investment advisers of 'soft dollar' benefits. In the absence of a safe harbour, such benefits would violate investment advisers' fiduciary duties. Introduced in 1975, section 28(e) provides a safe harbour permitting investment advisers to pay above-market commissions to receive particular benefits from the broker-dealers to which they direct client orders for execution.[121] The safe harbour applies only if investment advisers determine in good faith that commissions attributable to the benefits are reasonable in relation to those benefits, and they generally disclose those benefits to clients.[122] Though 'soft dollar' benefits represent one instance of the trilateral dilemma, their regulatory treatment reflects the type of piecemeal approach common in US COB regulation.

A further area where remuneration risks may skew advice relates to research analysts, who work for broker-dealer firms. In the wake of scandals following the late 1990s market boom, when research analysts were shown to have skewed their equity research,[123] the SEC, FINRA, and the NYSE implemented rule changes buttressing analyst 'independence'. Among other things, these regulations attempted to insulate research analysts from pressure applied by investment bankers and

[119] eg, Ferrell, A, 'A Proposal for Solving the "Payment for Order Flow" Problem' (2001) 74 *Southern California Law Review* 1027; Jackson, H and Burlingame, L, 'Kickbacks or Compensation: The Case of Yield Spread Premiums' (2007) 12 *Stanford Journal of Law, Business & Finance* 289; Johnsen, D, 'Property Rights to Investment Research: The Agency Costs of Soft Dollar Brokerage' (1994) 11 *Yale Journal on Regulation* 75.

[120] Jackson, n 118 above, 82, 100.

[121] Although expressed to apply to persons who exercise 'investment discretion' over clients' accounts, section 28(e) applies primarily to investment advisers. See SEC, 'Guidance Regarding Client Commission Practices under section 28(e) of the Securities Exchange Act of 1934' (2006), 71 *Federal Register* 41,978–42,051, 41,978 n 3.

[122] See n 40 above, section 28(e).

[123] For a survey of the empirical evidence, see Mehran, H and Stulz, R, 'The Economics of Conflicts of Interest in Financial Institutions' (2007) 85 *Journal of Financial Economics* 267.

required broker-dealer firms to make aggregate disclosures to highlight their incentives to promote trading activity.[124]

5. Regulatory oversight

A recent assessment of the regulatory oversight shows significant differences between broker-dealers and investment advisers.[125] Broker-dealers are subject to more compliance examinations and enforcement actions than investment advisers. According to the study, the number of SEC examinations of investment advisers conducted annually 'decreased 29.8 per cent, from 1,543 examinations in 2004 to 1,083 examinations in 2010'.[126] The fall has been attributed to the growth in the number of investment advisers and their assets under management.[127] The study observed that in 2010 only 9 per cent of investment advisers were examined by the SEC, while over 50 per cent of broker-dealers were examined by FINRA.[128]

IV. International Comparisons

While COB regulation does not benefit from international standard setting, the regulatory strategies employed in important jurisdictions are remarkably similar. Both the EU and Australian regimes require providers of financial services, including advisory and execution services, to be registered.[129] Standards of care, loyalty, and fair dealing must be met,[130] and duties of best execution are also owed.[131] The

[124] SEC, Regulation Analyst Certification; FINRA Manual, Rule 2711; NYSE Rule 472. Transgressions by research analysts led to enforcement action resulting in a 'Global Settlement' involving major financial institutions. See n 43 above.

[125] SEC, Enhancing Investment Adviser Examinations, n 71 above. [126] ibid, 14.

[127] Walter, E, Statement on Study on Enhancing Investment Adviser Examinations (2011), available at <https://www.sec.gov/news/speech/2011/spch011911ebw.pdf> (last accessed 23 June 2014).

[128] SEC, Enhancing Investment Adviser Examinations, n 71 above, 14, 30–1.

[129] Under MiFID I, firms must be authorized to provide investment services. MiFID I Level 1 Directive, Article 5. For an extensive discussion, see Moloney, N, *EC Securities Regulation* (2nd edn, 2008) 410–23. Australia requires businesses (but not individuals) providing financial services—including those who provide advice about a financial product—to be licensed. Corporations Law (Cth), section 911A.

[130] As to fair dealing, see MiFID I Level 1, Article 19(1) and Corporations Act 2001 (Cth), section 912A(1)(a). Standards of care and loyalty are considered below.

[131] See MiFID I Level 1, Articles 21 and 22, and MiFID I Level 2, Article 48 and ASIC, Market Integrity Rules (Competition in Exchange Markets) (2011).

EU duties of care (and suitability in particular), fair dealing, and best execution are broadly similar to those in the US.[132]

Some important differences nevertheless exist among regimes. For instance, the EU clearly distinguishes among categories of clients and is less willing to regard disclosure as sufficient in managing conflicts of interest. The precise scope of the primary legal instruments varies between jurisdictions.[133] One further difference concerns the extent to which regimes are prepared to undertake major reforms. The US remains tied to the regulatory structure adopted in the mid-1930s and its changes have been incremental and piecemeal. The EU and Australia, in contrast, have undertaken wholesale reforms of COB regulation, perhaps reflecting the less adversarial and contested nature of the politics of financial regulatory reform in these jurisdictions.

1. Standards of conduct

(a) *Client categorization*

MiFID I adopts a threefold categorization of clients, according to their knowledge and experience. It provides them with differing levels of protection, except in the area of conflicts of interest, where all categories receive equal protection. Professional clients are those that meet certain qualitative thresholds regarding their experience, knowledge, and expertise, and that fall into any of a number of enumerated investor classes.[134] For some purposes, professional clients may be eligible counterparties. The residual category is retail clients. Investment firms must inform clients of their categorization and give each the right to request a different categorization. Importantly, eligible counterparties do not receive some protections; in particular, they are not owed duties as to suitability, appropriateness, best execution, order handling, or inducements.[135] Professional clients are owed a somewhat diluted suitability duty. In contrast, US regulation does not categorize clients in the same way, but does prevent some clients from investing in certain securities and relaxes the suitability obligation owed to institutional clients.[136]

[132] The World Bank, Comparing European and US Securities Regulations: MiFID versus Corresponding US Regulations (2010), World Bank Working Paper No 184, 22 (noting the similar duties, but observing that the EU's best execution duty places less emphasis on price than the corresponding US duty).

[133] For instance, MiFID I is both narrower and broader than COB regulation relating to investment advisers and broker-dealers. It is narrower in that it does not apply to advisers to non-discretionary accounts; it is broader in that it applies to financial instruments in addition to securities, such as swaps and futures. See DeMott and Laby, n 8 above, 412–14.

[134] MiFID I Level 1, Annex II.

[135] For discussion, see Moloney, n 129 above, 591–623; and Nelson, n 8 above, 225–53.

[136] See n 138 below and accompanying text.

(b) *Care*

Under MiFID I, the EU imposes a process-based suitability duty on investment firms providing investment advice or portfolio-management services. For retail clients, the duty requires an investment firm to obtain the information 'necessary' for it to have a 'reasonable basis for believing' that its recommendation is suitable. Suitability is defined in terms of the client's investment objectives, risk-bearing ability, and experience and knowledge that would enable it to understand the risks involved.[137] Like the equivalent US rule, the duty requires financial intermediaries to conduct investigations of their clients to establish a 'reasonable basis' for believing that the advice or recommendation is suitable for the client, given the client's characteristics. Neither rule requires that the advice or recommendation (that is, the outcome of that process) in fact be suitable.

MiFID I's suitability duty applies differentially, depending on the category of client involved. The duty set out above applies to retail clients. In the case of professional clients, investment firms may assume that two of the suitability-related factors are satisfied, namely that the clients have the financial capacity to bear the investment risks involved and have the knowledge and experience necessary to understand those risks. In contrast, the US regime dilutes broker-dealers' suitability obligation for 'institutional' clients (a category corresponding to professional clients) where firms have a 'reasonable basis to believe that the institutional [client] is capable of evaluating investment risks independently ...'.[138]

MiFID I also imposes a less-protective 'appropriateness' duty for non-advised services, such as execution and transmission services.[139] No equivalent duty applies in the US. The appropriateness duty requires an assessment of the investor's knowledge and experience, but not of her financial situation or objectives. The duty does not apply for sales of 'non-complex' products.

The Australian regime, which came into force in 2013, provides an interesting counterpoint to the EU and US approaches. It imposes a heightened suitability duty, requiring that advice be both in the 'best interests' of a client and 'appropriate' for it.[140] The regime replaced an outcome-based appropriateness duty (which required financial intermediaries to ensure their advice was appropriate for clients) after a 2009 Parliamentary Joint Committee review considered the duty too weak.[141] The duties apply to the giving of personal financial product advice to retail clients. The statute does not define the 'best interests' standard,

[137] MiFID I Level 1, Article 19(4) and MiFID I Level 2, Article 35(1).

[138] FINRA Manual, Rule 2111(b).

[139] MiFID I Level 1, Article 19(5).

[140] Corporations Act 2001 (Cth), sections 961B and 961G. The duties apply to individual advisers, rather than to firms.

[141] Commonwealth of Australia, Parliamentary Joint Committee on Corporations and Financial Services: Inquiry into Financial Products and Services in Australia (2009), 87.

but does prescribe seven 'steps' that, if taken, will allow a 'provider' of advice to satisfy the duty.[142] The statute further requires that it be 'reasonable to conclude' that resulting advice is 'appropriate' to the client.[143] The Australian Securities and Investment Commission will judge 'appropriateness' from the perspective of an advice provider who has complied with the 'best interests' duty.[144] These duties may not be contractually displaced or varied.[145] No equivalent duty applies to protect institutional investors.

(c) *Loyalty*

Although the EU and Australia, like the US, impose 'fiduciary' or 'best interest' duties on advisers, the relevant duties all differ from the conflict-avoidance standard imposed under the general law's fiduciary doctrine.[146] Under MiFID I, the 'fair treatment' obligation in Article 19(1) requires an investment firm to 'act honestly, fairly and professionally in accordance with the best interests of its clients …'; an obligation the European Commission described as 'reinforced fiduciary duties'.[147] The specific MiFID I conflict of interest provisions emphasize reasonableness and focus more on the intended effect or design of preventative measures than on their practical effectiveness. Article 13(3) of MiFID I requires investment firms to '[take] all reasonable steps', using organizational and other arrangements, 'designed to prevent conflicts of interest … from adversely affecting the interests of its client'.[148] Despite the reference to 'prevent[ing] conflicts of interest', MiFID I makes clear the obligation is in fact to 'manage' conflicts of interest.[149] Moreover, Article 18(2) suggests that, even where organizational and other arrangements are insufficient to prevent client harm, disclosure—rather than informed client consent—will allow the financial intermediary to proceed.[150] The rules regulating conflicts of interest do not vary according to the category of client.

[142] See n 140 above, section 961B(2). ASIC, Licensing: Financial Product Advisers—Conduct and Disclosure, Regulatory Guide 175 (2013), 67–8. According to ASIC, other steps may also satisfy the best interests duty provided they 'produce at least the same standard of advice for the client' as the safe harbour. See ASIC, 65.

[143] See n 140 above, section 961G. [144] ASIC, n 142 above, 85–6.

[145] ibid, 85. [146] See Section II.1.

[147] European Commission, Retail Financial Services Green Paper, 12; Moloney, n 4 above, 213, n 132.

[148] MiFID I Level 1, Article 13(3); MiFID I Level 2, Recital 27.

[149] Article 18(2) refers to the Article 13(3) duty as one to 'manage' conflicts of interest, and firms are required to have in place policies that identify and then 'manage' conflicts of interest. MiFID I Level 2 Directive, Article 22.

[150] In those circumstances, MiFID I Level 1, Article 18(2) requires firm must 'clearly disclose the general nature and/or the sources of the conflicts of interest to the client before undertaking business on its behalf'. See also MiFID I Level 2, Articles 21–3. MiFID I Level 2, Article 22(4) elaborates on the Article 18(2) duty, but falls short of requiring the client's informed consent. See Moloney, n 4 above, 213–14 (suggesting limits on the extent to which disclosure and investor consent can satisfy these duties).

In its implementation of MiFID I, the UK takes an investor-protection view of MiFID I's conflict-management obligation. In its guidance, the Financial Services Authority (FSA) (the Financial Conduct Authority's predecessor) referred to the 'fair treatment' obligation in acknowledging that disclosure may be ineffective to manage some conflicts of interest.[151] Instead, the FSA explained, 'a firm should identify the actual and potential conflicts of interest, and put in place effective arrangements to mitigate those risks'.[152] Additionally, '[i]f a firm cannot manage a conflict, it must carefully consider whether it would be in the best interests of [its] client to go ahead with a transaction or service'.[153] MiFID I's conflict of interest rules fall short of a 'best interest' requirement incorporated into the Article 19(1) 'fair treatment' obligation, and yet are consistent with the US approach of allowing financial intermediaries considerable discretion in determining how to address conflicts of interest.

Australia's conflict of interest regime is noteworthy because it is stricter again than those of the US and the EU. In addition to requiring firms to 'manage' conflicts of interest,[154] the regime introduced a 'conflict priority duty' in its 2013 reforms.[155] The regime also applies beyond securities to banking and insurance activities. The conflicts priority duty requires an advice provider to prioritize the interests of her client where she 'knows, or reasonably ought to know, that the client's interests conflict with [her] own or those of a related party'.[156] Importantly, the duty cannot be discharged or excluded by either client consent or use of a contractual disclaimer.[157]

Generalized comparisons with the US are difficult. Nevertheless, MiFID I imposes a broad-based conflict-'management' obligation, while the US relies more on disclosure and, in some contexts, specific bans.[158] Both regimes fall short

[151] Financial Services Authority, Platforms and More Principles-based Regulation, Feedback Statement 08/01 (2008) 19. The FSA disagreed with the notion that disclosure could cure all conflicts of interest, referring to principle 8 that a firm must 'manage conflicts of interest fairly, both between itself and its clients and between a client and another client'.

[152] ibid, 19. [153] ibid.

[154] See n 140 above, section 912A(1)(aa) (requiring 'adequate arrangements for the management of conflicts of interest'). As to its meaning, see *ASIC v Citigroup Global Markets Australia Pty Ltd* (No 4) (2007) 160 FCR 35 [444]–[445]. cf ASIC, Licensing: Managing Conflicts of Interest, Regulatory Guide 181 (2004) and Tuch, A, 'The Paradox of Financial Services Regulation: Preserving Client Expectations of Loyalty in an Industry Rife with Conflicts of Interest' in Tjio, H (ed.), *The Regulation of Wealth Management* (2008) 53.

[155] See n 140 above, section 961J.

[156] Importantly, the 'conflict priority duty' does not attribute all information in possession of the firm to the adviser in question. According to ASIC, an individual adviser (to whom the duty applies) will be taken to know conflicts of interest disclosed by her firm in its financial services guides, but otherwise will not make inquiries as to the interests of related parties. See ASIC, n 142 above, 91.

[157] ibid.

[158] The World Bank, Comparing European and US Securities Regulations: MiFID versus Corresponding US Regulations (2010), World Bank Working Paper No 184, 21 (describing MiFID I's conflict of interest rules as 'broad and general' and those in the US as 'focused on specific situations').

of requiring financial intermediaries to avoid conflicts of interest in the absence of informed client consent and both give financial intermediaries considerable discretion in addressing conflicts of interest. Australia's regime would seem more client-protective, although its effectiveness is still to be tested.

2. Remuneration-based risks

Like the US, the EU relies on its process-based suitability duty as well as its conflict of interest rules to contain the risks of commission-based compensation. As discussed above, MiFID I's conflict of interest rules rely heavily on disclosure—a strategy that may be poorly suited to protecting trusting retail investors. According to Professor Niamh Moloney, MiFID I 'does not appear to be notably successful in addressing the most acute retail market risks concerning commissions in the sales and advice process'.[159]

The EU more successfully combats the trilateral dilemma, which arises where financial intermediaries receive third-party benefits that may skew the independence of their advice. MiFID I limits investment firms' receipt of inducements in giving advice to or exercising discretion on behalf of clients. MiFID I bans any benefits (including fees, commissions, and non-monetary benefits) in connection with the provision of investment services and activities, namely investment advice, portfolio management, and trade execution for clients.[160] MiFID I permits such benefits provided to or by a third party, provided the benefits are disclosed; are 'designed to enhance the quality of the relevant service to the client'; and do not 'impair' the firm's duty (under Article 19(1)) to act in the client's best interests.[161] The regime thus accepts the potential desirability of third-party benefits, but also imposes broad constraints. The EU once again imposes general rules where the US adopts a piecemeal, context-specific approach.

Australia's recently implemented inducement regime is more strict. It bans financial intermediaries from accepting certain types of remuneration considered to materially sway their ability to give financial product advice to retail clients.[162] The statute prohibits financial services licensees, and their representatives, from accepting 'conflicted remuneration' in providing financial product advice to a retail client.[163] The statute broadly defines 'conflicted remuneration' as any benefit, monetary or non-monetary, the nature or circumstances of which 'could reasonably be

[159] Moloney, n 4 above, 247. According to Professor Moloney, MiFID I relies in part on disclosure to manage commission risks. Moloney, n 4 above, 263.

[160] Level II Directive, Article 26 bans these other inducements by rendering them in violation of the Article 19(1) duty to act honestly, fairly, and professionally in accordance with the best interests of a client.

[161] Level II Directive, Article 26(b). With minor exceptions, the provision bans inducements (as violating the Article 19(1) fair-treatment obligation) unless these conditions are met.

[162] See n 140 above, Div 4 of Part 7.7A.

[163] ibid, sections 963E and 963A.

expected to influence' either the licensee's or the representative's financial product advice or the 'financial product recommended'.[164] It is not evident what changes these reforms have made to market practices and their desirability is also open to question. They implicitly deny the possibility that third-party benefits may improve the quality of a firm's advice or reduce its costs—heavily contested propositions.[165]

3. Enforcement and effectiveness

Generalizations about a regime's intensity of enforcement are difficult to make. Nevertheless, US broker-dealers are more robustly regulated than investment advisers. As between jurisdictions, the US maintains a level of enforcement staff exceeding that of most other countries, even taking account of market size. Private enforcement through the class-action device provides significantly greater deterrent force in the US than elsewhere,[166] although the common use of mandatory arbitration clauses by broker-dealers and investment advisers limits the force of this device. Where litigation does arise, the US pretrial discovery system provides additional deterrence. The deterrent force of COB regulation will vary from country to country within the EU due to differences in enforcement apparatus. The notion that US enforcement was more intensive than UK enforcement, in particular, prevailed in the lead-up to the global financial crisis of 2007–09, with the UK regarded as employing a 'light touch' approach. Nevertheless, since optimal deterrence is so difficult to assess in cross-country comparisons,[167] little would be gained by drawing conclusions.

V. Financial Crisis and Other Recent Developments

While reforms to promote financial stability have been the focus of post-financial crisis regulatory developments, COB regulation has also been tightened. Australia's regime incorporates post-global financial crisis reforms, although the EU and the US have yet to introduce many crisis-inspired reforms. The Dodd–Frank Act tasked regulators with studying various issues and implementing their recommendations. One is the long-standing concern (predating the financial crisis) about the distinct regulatory regimes—and different rules of conduct—for broker-dealers

[164] ibid, section 963A. [165] See nn 118 and 119 above. [166] Cox, n 9 above, 104–5.

[167] Jackson, H, 'Variation in the Intensity of Financial Regulation: Preliminary Evidence and Potential Implications' (2007) 24 *Yale Journal on Regulation* 101.

and investment advisers, based on an increasingly blurred distinction that confuses clients.[168] Tasked with assessing the desirability of imposing a fiduciary duty on broker-dealers when they provide personalized investment advice to retail clients,[169] the staff of the SEC recommended that a 'uniform fiduciary standard' be adopted for both broker-dealers and investment advisers.[170] It seems clear that the duty would lack an implied private right of action. According to the SEC, the proposed duty would oblige investment advisers and broker-dealers to 'eliminate or disclose conflicts of interest'; and further, that 'certain' (unidentified) conflicts would be prohibited.[171] The SEC has yet to implement its recommendation, but one must doubt whether the reforms will have much impact beyond intermediaries' disclosure practices, and little suggests they will combat broker-dealers' commission-based incentives. The Dodd–Frank Act also empowers the SEC to promulgate rules to prohibit or restrict compensation schemes for broker-dealers,[172] but to date nothing suggests the SEC will do so. Similarly, the SEC has indicated no willingness to prohibit or restrict the use of mandatory pre-dispute arbitration clauses by broker-dealers and investment advisers, despite having been granted such power in the Dodd–Frank Act.[173] However, the US COB regime will continue along the path of increasing complexity—the proposed fiduciary standard will 'overlap on top of the existing investment adviser and broker-dealer regimes', adding a further layer of COB regulation.[174]

As a result of another study mandated by the Dodd–Frank Act,[175] the SEC demonstrated the inferior examination and enforcement resources for investment advisers relative to broker-dealers. It predicted that the SEC 'will not have sufficient capacity in the near or long term to conduct effective examinations of registered investment advisers with adequate frequency'.[176] The SEC referred the issue to Congress, suggesting that Congress either levy fees on SEC-registered investment advisers to fund examinations, adopt self-regulatory oversight for investment advisers, or authorize FINRA to examine those investment advisers that are dual-registered as broker-dealers. Congress has yet to act.

Other financial crisis-related COB reforms provide increasing protections for institutional investors. The notion of institutional clients as able to 'fend for themselves' came under attack most dramatically with the SEC enforcement action against Goldman Sachs in 2010 for its sale to sophisticated clients of collateralized debt obligations that were 'designed to fail'. Given the moniker ABACUS 2007-AC1,

[168] As to client confusion, see SEC, n 10 above, 51, 101; and Laby, A, 'Selling Advice and Creating Expectations: Why Brokers Should Be Fiduciaries' (2012) *Washington Law Review* 707, 736–9.

[169] Dodd–Frank Act, section 913. [170] SEC, n 10 above, 51, 101, and 103.

[171] ibid, vi–vii. [172] See n 169 above, section 913(g).

[173] ibid, section 921. See also SEC, n 10 above, 44 and 79.

[174] SEC, n 10 above, 109.

[175] SEC, Enhancing Investment Adviser Examinations, n 71 above.

[176] ibid, 3–4.

the transaction led to losses for institutional investors of around $1 billion and produced calls, from Congress and elsewhere, for the imposition of fiduciary duties on broker-dealers, including in their dealings with institutions.[177] The enforcement action gave impetus to Congress' adoption of the Volcker Rule, which bans financial intermediaries with bank affiliates from engaging in a broad range of trading-related activities, including trading on behalf of clients, if doing so would give rise to a 'material conflict of interest'.[178] The implementing regulations provide that a conflict of interest will not be 'material' where an intermediary has disclosed it or used information barriers, unless the intermediary 'knows or should reasonably know that ... [nevertheless] the conflict of interest may involve or result in a materially adverse effect on a client, customer, or counterparty'.[179] The Dodd–Frank Act also created two new categories of market participants in derivatives markets—swap dealers and major swap participants—and imposed new COB standards on them, including a 'best interests' duty when advising a state, municipality, pension plan, or endowment.[180]

The 2014 MiFID II/MiFIR regime will retain the central pillars of existing EU COB regulation, but will significantly bolster the regulation of remuneration-based risks. Investment firms will have to inform clients and potential clients 'in good time' whether their advice is 'provided on an independent basis'.[181] Where it is, the firm will be forbidden from accepting fees, commissions, or other benefits from a third party in relation to the service to clients.[182] The regime will thus require clients to pay fees for 'independent' advice and mirror recent reforms in Australia, the UK, and the Netherlands. While likely to significantly diminish the exploitation of remuneration-based conflicts of interest, the prohibition will do nothing for those clients who do not receive independent advice, such as those potentially investing in proprietary products offered by firms. Other provisions in the MiFID II/MiFIR regime will oblige investment firms to ensure they do not remunerate or evaluate staff 'in a way that conflicts with [their duties] to act in the best interests of [their] clients'.[183] Another change will require investment firms that 'manufacture financial instruments for sale to clients' to ensure those instruments are designed to meet the needs of relevant clients[184]—a provision apparently responsive to Goldman's activities in the ABACUS transaction.

The 2014 MiFID II/MiFIR regime will retain the suitability and appropriateness duties, with some additions. When providing investment advice to a retail client,

177 McKinnon, J, 'Lawmakers Target Investment Banks' *Wall Street Journal Online*, 5 May 2010. See also Tuch, n 26 above, 368–70.

178 The Dodd–Frank Act amended the Bank Holding Company Act of 1956 by introducing a new section 13.

179 See 17 CFR 255.7(b)(2)(ii).

180 See n 169 above, Title VII.

181 2014 MiFID II, Article 24(4).

182 ibid, Article 24(7).

183 ibid, Article 24(10).

184 ibid, Article 24(2).

investment firms will have to provide a statement to the client specifying the advice given and how it 'meets the preferences, objectives and other characteristics' of the client.[185] When bundling products and services together, an investment firm will need to apportion the costs of each component, inform clients whether the different components may be bought separately, and even inform clients when bundling creates risks different from those of the component parts.[186] The regime will also provide greater protection for clients trading complex products, by amending the scope of application of the appropriateness duty. While non-complex products will remain outside the rule's reach, structured undertakings for collective investment in transferable securities (UCITS) will now be regarded as complex and thus within the rule's scope.[187] Proposed changes to the client categorization regime, though seemingly minor, include imposing a standard of fair dealing on investment firms in their relationships with eligible counterparties.[188]

Regulators are also responding to pressures arising from fragmented trading markets and new technologies. In the US, trading in exchange-listed equities occurs in multiple trading venues, including 11 exchanges, more than 40 alternative trading systems, and hundreds of broker-dealers.[189] While the competition among venues may lower some trading costs, it also affords broker-dealers more options in executing trades. Because some venues pay broker-dealers for order flow (an instance of the trilateral dilemma) and clients are unable adequately to police broker-dealers' execution decisions, the potential for client harm exists. These risks to client loyalty are exacerbated by technological developments, which include the increasing use of algorithmic trading strategies. The SEC is focusing on these risks and incremental reform to COB regulation can be expected.[190] One proposal involves narrowing the exemptions high-frequency trading firms may rely on to avoid broker-dealer and FINRA registration.[191]

VI. Conclusion

COB regulation in the US is characterized by complexity, piecemeal reform, and a blunt distinction between financial intermediaries that is not adopted in comparable jurisdictions. Unlike other jurisdictions, particularly the EU and Australia,

[185] ibid, Article 25(6). [186] ibid, Article 24(11).
[187] ibid, Article 25(3), (4). [188] ibid, Article 30(1).
[189] White, n 20 above. For a description of the risks to investors arising from changes to market structure and technological developments, see Lewis, M, *Flash Boys* (2014).
[190] White, n 20 above. [191] ibid.

the US has also shown resistance to addressing remuneration-based risks, especially commission risks facing broker-dealers. Still, whether US COB regulation produces weaker deterrence than other regimes is difficult to tell. More detailed cross-jurisdictional analysis would be desirable, particularly regarding liability that arises from breaches of obligations imposed by COB regulation. It is, nevertheless, apparent that US enforcement is robust, at least for broker-dealer interactions with retail clients. While elaborate and often esoteric, the US regime may even be more tailored to subtle differences in financial intermediaries, products, and markets than other regimes. If that were so, however, it would be more the result of good fortune than of careful design.

Bibliography

Bank for International Settlements, The Joint Forum. Client Suitability in the Retail Sale of Financial Products and Services (2008).

Baxt, R, Black, A, and Hanrahan, P, *Securities and Financial Services Law* (7th edn, 2008).

Birdthistle, W and Henderson, T, 'Becoming the Fifth Branch' (2013) 99 *Cornell Law Review* 1.

Black, B, 'Brokers and Advisers—What's in a Name?' (2005) 11 *Fordham Journal of Corporate and Financial Law* 31.

Black, B, 'How to Improve Retail Investor Protection after the Dodd-Frank Wall Street Reform and Consumer Protection Act' (2010) 13 *University of Pennsylvania Journal of Business Law* 101.

Black J, 'Law and Regulation: The Case of Finance' in Braithwaite, J et al. (eds), *Regulating Law* (2004) 33.

Campbell, J, Jackson, H, Madrian, B, and Tufano, P, 'Consumer Financial Protection' (2011) 25 *Journal of Economic Perspectives* 9.

Campbell, J, Jackson, H, Madrian, B, and Tufano, P, The Regulation of Consumer Financial Products: An Introductory Essay with Four Case Studies (2010), Harvard Kennedy School Faculty Research Working Paper Series, available at <http://dash.harvard.edu/bitstream/handle/1/4450128/Madrian_TheRegulationof.pdf?sequence=1>, 7–11.

Carnell, R, Macey, J, and Miller, G, *The Law of Financial Institutions* (5th edn, 2013).

Coffee, J and Sale, H, 'Redesigning the SEC: Does the Treasury Have a Better Idea?' (2009) 95 *Virginia Law Review* 707.

Coffee, J and Sale, H, *Securities Regulation: Cases and Materials* (12th edn, 2012).

Commonwealth of Australia, Parliamentary Joint Committee on Corporations and Financial Services: Inquiry into Financial Products and Services in Australia (2009).

Cox, J, Hillman, R, and Langevoort, D, *Securities Regulation: Cases and Materials* (7th edn, 2013).

DeMott, D, 'Beyond Metaphor: An Analysis of Fiduciary Obligation' (1988) 37 *Duke Law Journal* 879.

DeMott, D and Laby, A, 'The United States of America' in Busch, D and DeMott, D (eds), *Liability of Asset Managers* (2012) 411.

Ferrell, A, 'A Proposal for Solving the "Payment for Order Flow" Problem' (2001) 74 *Southern California Law Review* 1027.
Gadinis, S, 'The SEC and the Financial Industry: Evidence from Enforcement Against Broker-Dealers' (2012) 67 *The Business Lawyer* 679.
Gray, J, 'London Money Markets: Legal and Regulatory Framework' in Blair, M et al. (eds), *Financial Markets and Exchanges Law* (2nd edn, 2012) 243.
Hazen, T, 'Are Existing Stock Broker Standards Sufficient? Principles, Rules, and Fiduciary Duties' (2010) *Columbia Business Law Review* 710.
Hudson, A, *The Law of Finance* (2nd edn, 2013).
Jackson, H, 'Regulation in a Multi-Sectored Financial Services Industry: An Exploratory Essay' (1999) 77 *Washington University Law Quarterly* 319.
Jackson, H, 'The Trilateral Dilemma in Financial Regulation' in Lusardi, AM (ed.), *Improving the Effectiveness of Financial Education and Savings* (2008).
Jackson, H and Burlingame, L, 'Kickbacks or Compensation: The Case of Yield Spread Premiums' (2007) 12 *Stanford Journal of Law, Business & Finance* 289.
Johnsen, D, 'Property Rights to Investment Research: The Agency Costs of Soft Dollar Brokerage' (1994) 11 *Yale Journal on Regulation* 75.
Karmel, R, 'Is the Shingle Theory Dead?' (1995) 52 *Washington and Lee Law Review* 1271.
Kirsch, C, *Investment Adviser Regulation: A Step-by-Step Guide to Compliance and the Law* (3rd edn, 2011).
Laby, A, 'Reforming the Regulation of Broker-Dealers and Investment Advisers' (2010) 65 *The Business Lawyer* 395.
Laby, A, 'SEC v. Capital Gains Research Bureau and the Investment Advisers Act of 1940' (2011) *Boston University Law Review* 1051.
Laby, A, 'Selling Advice and Creating Expectations: Why Brokers Should Be Fiduciaries' (2012) *Washington Law Review* 707.
Langevoort, D, 'Brokers as Fiduciaries' (2010) 71 *University of Pittsburgh Law Review* 439.
Langevoort, D, 'Selling Hope, Selling Risk: Some Lessons for Law from Behavioral Economics about Stockbrokers and Sophisticated Clients' (1996) 84 *California Law Review* 627.
Langevoort, D, 'Taming the Animal Spirits of the Stock Markets: A Behavioral Approach to Securities Regulation' (2002) 97 *Northwestern University Law Review* 135.
MacNeil, I, *An Introduction to the Law on Financial Investment* (2012).
Markham, J, 'Protecting the Institutional Investor—Jungle Predator or Shorn Lamb' (1995) 12 *Yale Journal on Regulation* 345.
Moloney, N, *EC Securities Regulation* (2nd edn, 2008).
Moloney, N, *How to Protect Investors: Lessons from the EC and the UK* (2010).
Nelson, P, *Capital Markets Law and Compliance: The Implications of MiFID* (2008).
Poser, N, 'Liability of Broker-Dealers for Unsuitable Recommendations to Institutional Investors' (2001) *Brigham Young University Law Review* 1493.
Poser, N and Fanto, J, *Broker-Dealer Law and Regulation* (4th edn, 2014).
Prentice, R, 'Whither Securities Regulation' (2002) 51 *Duke Law Journal* 1397.
Schwarcz, D, 'Beyond Disclosure: The Case for Banning Contingent Commissions' (2007) 25 *Yale Law & Policy Review* 289.
SEC, Staff Study on Enhancing Investment Adviser Examinations as Required by Section 914 of the Dodd–Frank Wall Street Reform and Consumer Protection Act (2011).

SEC, Study on Investment Advisers and Broker-Dealers, as Required by Section 913 of the Dodd–Frank Wall Street Reform and Consumer Protection Act (2011).

Smethurst, J, Heukamp, W, Goodliffe, J, and Miller, R et al., 'Conduct of Business Regulation: A Survey of the UK Regime and a Comparison with the US, German and Hong Kong Approach' in Burling, J and Lazarus, K (eds), *Research Handbook on International Insurance Law and Regulation* (2011) 355.

Tuch, A, 'Conflicted Gatekeepers: The Volcker Rule and Goldman Sachs' (2012) 7 *Virginia Law and Business Review* 365.

Tuch, A, 'Financial Conglomerates and Information Barriers' (2014) 39 *Journal of Corporation Law* 563.

Tuch, A, 'The Paradox of Financial Services Regulation: Preserving Client Expectations of Loyalty in an Industry Rife with Conflicts of Interest' in Tjio, H (ed.), *The Regulation of Wealth Management* (2008) 53.

Wrona, J, 'The Best of Both Worlds: A Fact-based Analysis of the Legal Obligations of Investment Advisers and Broker-Dealers and a Framework for Enhanced Investor Protection' (2012) 68 *The Business Lawyer* 1.

CESifo WORKING PAPERS

7275
2018
September 2018

Labor Responses, Regulation and Business Churn

Marta Aloi, Huw Dixon, Anthony Savagar

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7275*
Category 6: Fiscal Policy, Macroeconomics and Growth

Labor Responses, Regulation and Business Churn

Abstract

We develop a model of sluggish firm entry to explain short-run labor responses to technology shocks. We show that the labor response to technology and its persistence depend on the degree of returns to labor and the rate of firm entry. Existing empirical results support our theory based on short-run labor responses across US industries. We derive closed-form transition paths that show the result occurs because labor adjusts instantaneously whilst firms are sluggish, and closed-form eigenvalues show that stricter entry regulation results in slower convergence to steady state.

JEL-Codes: D250, E200, L110, O330.

Keywords: deregulation, dynamic entry, endogenous entry costs.

Marta Aloi
University of Nottingham
Nottingham / United Kingdom
marta.aloi@nottingham.ac.uk

Huw Dixon
Cardiff University
Cardiff / United Kingdom
dixonh@cardiff.ac.uk

Anthony Savagar
University of Kent
Canterbury / United Kingdom
a.savagar@kent.ac.uk

September 18, 2018
Thank you to Mathan Satchi, Christian Siegel, Mehdi Sahneh, Vincent Sterk, Petr Sedlacek, Kjetil Storesletten and Matteo Cacciatore for useful discussions.

The short-run response of labor hours to technology shocks is widely debated in macroeconomics.[1] Empirical studies, such as Chang and Hong 2006, document different labor responses to technology shocks across U.S. manufacturing industries. They show that while some industries exhibit a temporary reduction in employment in response to a permanent increase in technology, many more industries exhibit a short-run increase in both employment and hours per worker. However, the theory underlying these responses is not fully understood. In this paper, we identify a novel mechanism based on dynamic firm entry to explain short-run labor responses and subsequent persistence. Cross-industry data supports our theory. Additionally, we show that persistence of labor responses depends on firm sluggishness which regulation affects through endogenous entry costs.

Our mechanism focuses on endogenous variation in labor *per firm* which occurs when firm creation is sluggish but labor adjusts instantaneously. Endogenous variation in labor per firm is important for aggregate labor responses if the marginal product of labor (MPL) in a firm's production function is non-constant. For example, if a positive technology shock increases hours, but the stock of firms is fixed, hours per firm increase. With short-run increasing MPL, the rise in hours per firm increases MPL, increases wages and increases hours. Subsequent firm entry decreases hours per firm, decreases MPL, decreases wage, and decreases labor to its long-run level.[2] This channel is typically overlooked because either labor per firm is fixed or the MPL is constant so wages do not respond.

We develop a DGE small open economy (SOE) model in continuous time extended to include dynamic firm entry.[3] There is no capital, only labor, and there is an internationally traded bond with world interest rates equal to the household discount rate. Hence the household perfectly smooths utility, so consumption dynamics do not play a role, which allows a closed-form

[1]Cantore, Ferroni, and Leon-Ledesma 2017 provide a recent survey. The classic references are Gali 1999 for positive responses and Christiano, Eichenbaum, and Vigfusson 2003 for negative responses. See also Basu, Fernald, and Kimball 2006; Christiano, Eichenbaum, and Vigfusson 2004; Wang and Wen 2011; Rebei 2014.

[2]With decreasing MPL, the fall in hours per firm from entry, increases MPL, increases wages and increases labor to its long-run level.

[3]Sen and Turnovsky 1990; Mendoza 1991 are early papers in the SOE-RBC literature.

analysis of firm dynamics. Households can invest in new firms by paying an endogenous sunk entry cost. Once operational, firms compete under monopolistic competition and pay a fixed overhead cost each period. The restriction to one state variable (number of firms) keeps eigenvalues tractable, so we can study persistence and short-run versus long-run effects analytically. To model dynamic entry we assume that the entry costs depend on the flow of entry due to a congestion effect caused by red tape (Datta and Dixon 2002). Our model is parsimonious in order to derive general analytic results and qualitatively replicates key stylized facts.[4]

Related Literature: As mentioned at the start, the work of Chang and Hong 2006 provides evidence on the heterogeneity of short-run employment responses to technology shocks. Our work provides a new explanation for their findings based on labor returns to scale, and is broadly supported by their data. In relation to existing theoretical literature, we generalize the firm production function for increasing, decreasing or constant MPL and combine this with dynamic firm entry whilst maintaining tractability.[5] This distills the importance of dynamic firm entry, and contributes to growing evidence that dynamic (sluggish), rather than static, entry is crucial to understand business cycle dynamics. Bilbiie, Ghironi, and Melitz 2012 (BGM) is the seminal work in this literature. They show that dynamic entry and endogenous markups greatly improve RBC moment matching, and their modelling approach has been successfully adopted in quantitative DSGE exercises.[6] Our mechanism to achieve sluggish entry differs as it relies on endogenous sunk costs. This modelling choice pertains to tractable continuous time analysis, and allows us to study how deregulation can increase business churn and thus speed of adjustment following short-run responses. Lewis 2009 provides

[4]Procylical net entry which lags the cycle (Campbell 1998; Bergin, Feng, and Lin 2016); the existence of variable returns to scale in labor (Basu, Fernald, and Kimball 2006); the existence of monopoly power (De Loecker and Eeckhout 2017); procylical average firm scale (capacity utilization) which is contemporaneous with the cycle (Burnside and Eichenbaum 1996); procylical firm profits also contemporaneous with the cycle (Lewis 2009; Bilbiie, Ghironi, and Melitz 2012); countercylical labor share (Young 2004); and procylical measured productivity (Basu and Fernald 2001; Jaimovich and Floetotto 2008).

[5]Barseghyan and DiCecio 2016 study the relationship between returns to scale and entry in a perfectly competitive Hopenhayn model.

[6]Etro and Colciago 2010; Lewis and Poilly 2012; Lewis and Winkler 2017.

evidence on the importance of entry congestion in macroeconomic propagation.[7] Cantore, Ferroni, and Leon-Ledesma 2017 (Fig. 1, p.70) implies that short-run responses have reversed over the past century in the US from decreasing to increasing, and that the deviation now persists for longer. We explain that this could be caused by a decline in business churn.[8] Lastly, we show that entry effects on aggregate output are non-trivial with variable returns to scale in labor (MPL slope). This features is crucial to our understanding of transition, but also adds a new element to analyses of *optimal entry* by Etro and Colciago 2010; Bilbiie, Ghironi, and Melitz 2016 who focus on endogenous markups with constant MPL.

Roadmap: Section 1 outlines the household problem; Section 2 analyzes firm production and entry decisions; Section 3 summarizes equilibrium, solves for steady-state and solves for transition paths; Section 4 analyzes labor responses; Section 5 shows that deregulation speeds-up convergence.

1 Household

There is a small open economy, with a world capital market interest rate r equal to the discount rate ρ of the Ramsey household.[9]

$$r = \rho \tag{1}$$

[7] A number of recent papers have adopted entry adjustment costs (Lewis and Poilly 2012; Bergin and Lin 2012; Loualiche et al. 2014; Berentsen and Waller 2015; Poutineau and Vermandel 2015).

[8] This relates to recent literature on *'declining business dynamism'* (Decker et al. 2018) that links 'declines in the pace of business formation' to slower reallocation of resources.

[9] This so-called knife-edge condition is a widely-discussed model closing device (Turnovsky 2002; Oxborrow and Turnovsky 2017). Under perfect foresight, this will cause steady-state to depend on initial conditions (Uribe and Schmitt-Grohé 2017, Ch 2 & 3), so the deterministic steady-state is history dependent. Schmitt-Grohé and Uribe 2003 analyse techniques to induce stationarity for approximating equilibrium dynamics in stochastic models. Since our model is deterministic, non-stationarity is not an issue (Turnovsky 1997, Ch. 3).

We assume King-Plosser-Rebelo preferences with logarithmic consumption

$$U(C, 1-H) = \ln C - \frac{H^{1+\eta}}{1+\eta} \tag{2}$$

$\eta \in (0, \infty)$ is inverse Frisch elasticity of labor supply to wages.[10] The household earns income from three sources: supplying labor at wage w, receiving interest income from net foreign bonds rB and receiving profit income Π from owning firms. The household treats profit income as a lump sum payment. The household solves:

$$\max \int_0^{\infty} U(C, H) e^{-\rho t} dt \tag{3}$$

$$\text{subject to} \quad \dot{B} = rB + wH + \Pi - C \tag{4}$$

$$B(0) = B_0 \tag{5}$$

$$\text{where} \quad r = \rho \tag{1}$$

Given KPR preferences the optimal solutions satisfy

$$\dot{\lambda} = 0 \implies \lambda = \bar{\lambda} \tag{6}$$

$$\bar{C} = \frac{1}{\bar{\lambda}} \tag{7}$$

$$H(w, \lambda) = (\lambda w)^{\frac{1}{\eta}}, \quad \eta \in (0, \infty) \tag{8}$$

where we use bar notation for variables that are constant over time. For a given wage, labor supply H is increasing in λ. Frisch elasticity of supply measures the substitution effect of a change in the wage rate on labor supply $H_w \frac{w}{H} = \frac{1}{\eta}$.[11] The perfect capital markets assumption $r = \rho$ (implies constant

[10]We ignore indivisible labor $\eta = 0$. Additive separability $U_{CH} = 0$ is sufficient for our results to hold when there are increasing marginal costs (decreasing returns to labor). But we require KPR preferences for the decreasing and constant marginal cost cases.

[11]See the appendix for full derivation of first-order conditions. We rule out indivisible labor $\eta = 0$ which would imply $C = w$. If $r \neq \rho$ then no interior steady state exists. The trajectory of consumption will then be either increasing $r > \rho$ or decreasing $r < \rho$ through time. There are many discussions of *'closing devices'* (or *'stationarity-inducing devices'*) in the SOE literature, which are necessary because the exogenous world interest rate causes an incomplete market. See Seoane 2015 based on Mendoza 1991. Oxborrow

consumption $\dot{\lambda} = 0$) and additively separable utility $U_{CH} = 0$ simplify dynamics.[12] The result is that the only dynamics in the model will be a result of firm entry, which will affect wage. The advantage is to pinpoint the precise role of firm entry. λ is the marginal utility of consumption: high λ means low consumption and vice versa. Lastly, to ensure the private agent satisfies the intertemporal budget constraint, the transversality condition must hold

$$\lim_{t\to\infty} \lambda B e^{-rt} = 0 \tag{9}$$

Hence the solution to the problem is characterized by two boundary conditions (5), (9) and two ordinary differential equations (ODEs) $\dot{\lambda}, \dot{B}$ that solve to give trajectories $B(t), \lambda(t)\ \forall t$. Subsequently $\lambda(t)$ gives $C(t)$ and in turn $H(t)$ through the static conditions. However before solving we need to characterize the endogenous behaviour of w and Π in general equilibrium according to factor market equilibrium.

2 Firms: Technology, Entry and Exit

The aggregate consumption good C is either imported or produced domestically by a perfectly competitive industry with a constant returns production function using intermediate inputs which are monopolistically supplied. There is a continuum of possible intermediate products, $i \in [0, \infty)$. At instant t, there is a range of active products defined by $N(t) < \infty$ so that $i \in [0, N(t))$ are active and available, whilst $i > N(t)$ are inactive and not produced. Hence total domestic output Y is related to inputs y_i by the following technology

$$Y = N^{\varsigma - \frac{\theta}{\theta-1}} \left[\int_0^N y_i^{(\theta-1)/\theta} di \right]^{\theta/(\theta-1)} \tag{10}$$

and Turnovsky 2017 give overview and close the model using demography.

[12] Additive separability $u_{CH} = 0$ creates the simple relationship between consumption and marginal utility of consumption. The presence of a small open economy and perfect international capital markets $\rho = r$ implies the household can completely smooth its consumption so $\dot{\lambda} = 0 \implies \lambda = \bar{\lambda}$. Therefore together they imply the marginal utility of consumption is unchanging over time.

where $\theta > 1$ is the elasticity of substitution between products. The N^{ς} multiplier captures any variety effect. We assume $\varsigma = 1$ so no variety effect which implies an increase in the range of intermediates does not affect the unit cost function.[13] Treating the unit price of the consumption good as the numeraire, under monopolistic competition the demand for each available product i takes the constant elasticity form

$$y_i = p_i^{-\theta} \frac{Y}{N^{\varsigma}} \tag{11}$$

with corresponding price elasticity of demand $\varepsilon_{py} \equiv \frac{dp_i}{dy_i}\frac{p_i}{y_i}$ given by $\varepsilon_{py} = -\frac{1}{\theta}$. There is a continuum of potential firms, and each firm can produce one product. At time t, firm $i \in [0, N(t))$ has labor demand h_i to supply output y_i using the technology

$$y_i = \begin{cases} Ah_i^{\nu} - \phi, & \text{if } Ah_i^{\nu} > \phi, \\ 0 & \text{else,} \end{cases} \tag{12}$$

where $\nu > 0$ captures labor returns to scale (slope of MPL): $\nu < 1$ decreasing returns; $\nu = 1$ constant returns; $\nu > 1$ increasing returns. $\phi \geq 0$ is a fixed overhead cost denominated in output terms. A is a technology parameter. The fixed cost implies that labor returns to scale ν are not equivalent to

[13] A common case is $\varsigma = 0$ which leads to a variety effect, we want to remove this as it will create an additional mechanism adding to the main result we want to distill. Without removing love of variety, N will enter the labor market equilibrium condition, even with constant returns to scale.

overall returns to scale measured as average cost over marginal cost[14]

$$\frac{AC}{MC} = \nu(1 + s_\phi) \tag{14}$$

where $s_\phi \equiv \frac{\phi}{y}$ is the fixed cost share in output. The marginal product of labor and its slope are

$$\text{MPL} = \nu \frac{y + \phi}{h}, \tag{15}$$

$$\text{MPL}_h = (\nu - 1)\frac{\text{MPL}}{h}, \tag{16}$$

The MPL is always positive, but can be increasing $\nu > 1$, decreasing $\nu < 1$ or constant $\nu = 1$ in hours, corresponding to increasing, decreasing or constant returns to labor at the firm-level.[15]

[14]The cost function dual of our production function is $\text{TC} = \text{MC}\,\nu(y + \phi)$. This follows because factor prices equal their marginal revenue product, in the case for labor $w = \text{MR} \times \text{MPL}$. An optimizing firm produces where $MR = MC$, hence as labor is the sole input $\text{TC} = wh = \text{MC} \times \text{MPL} \times h = \text{MC}\,\nu(y+\phi)$. Multiply by $\frac{1}{y\,\text{MC}}$ to get AC / MC which captures overall returns to scale. Furthermore, where w is nominal wage, as labor is the only input, total costs are $\text{TC} = wh = w\left(\frac{y+\phi}{A}\right)^{\frac{1}{\nu}}$ so that marginal cost is

$$\text{MC} = \frac{w}{\nu A}\left(\frac{y + \phi}{A}\right)^{\frac{1-\nu}{\nu}} = \frac{\text{TC}}{\nu(y + \phi)} \tag{13}$$

and average cost is $\text{AC} = \frac{\text{TC}}{y}$ which in the U-shaped AC case ($\nu < 1$ and $\phi > 0$) will achieve minimum at firm scale $y^{\text{MES}} = \frac{\nu\phi}{1-\nu}$, the firm's *minimum efficient scale* (MES).

[15]When $\nu < 1$, $\phi > 0$ there is a U-shaped average cost curve with increasing marginal cost. This is compatible with both perfect and imperfect competition. When $\nu = 1, \phi = 0$, there are constant returns to scale: AC = MC. When $\nu = 1, \phi > 0$, there is a constant MC and decreasing AC. When $\nu > 1$ there is decreasing AC and MC. The extent of increasing returns to labor $\nu > 1$ is limited by the degree of imperfect competition. In the two cases with globally increasing returns to scale, equilibrium can only exist with imperfect competition.

2.1 Aggregate Output

Perfect factor markets imply aggregate labor is divided equally across firms $h_i = H/N,\ \forall i \in N$. Under symmetry the aggregate production function is

$$Y(N,H) = Ny = AH^{\nu}N^{(1-\nu)} - N\phi \tag{17}$$

It is homogeneous of degree 1 in inputs H, N which implies

$$Y = Y_N N + Y_H H \tag{18}$$

The intuition corresponds to $Y = Ny$. Output per firm is homogeneous of degree 0 because a change in aggregate labor is offset by a change in number of producers so that labor per firm is unchanged, then output per firm is unchanged, hence aggregate output expands proportionally to the expansion in number of firms. Treating N, H independently, the effect of entry on aggregate output is ambiguous whereas extra labor always raises aggregate output[16]

$$Y_N \equiv \frac{\partial Y}{\partial N} = (1-\nu)Ah^{\nu} - \phi = y - \nu Ah^{\nu} = (1-\nu)y - \nu\phi \gtreqless 0 \tag{19}$$

$$Y_H \equiv \frac{dY}{dH} = A\nu(H/N)^{\nu-1} = A\nu h^{\nu-1} = \nu\frac{y+\phi}{h} > 0 \tag{20}$$

When there are increasing returns to labor $\nu > 1$, an additional firm dividing aggregate labor into smaller units can decrease aggregate output as it employs labor less productively than the incumbents did prior to its entry. Aggregate and firm level MPL are equivalent $Y_H = y_h$.

[16] It is important to note the N derivative is partial, as the in general equilibrium the total derivative would recognize that a variation in N implicitly varies H, that is $\frac{dY}{dN} = \frac{\partial Y}{\partial N} + \frac{dY}{dH}\frac{dH}{dN}$. Since N is independent of H then its partial and total derivative are equivalent.

2.2 Profits and Factor Market Equilibrium

Due to imperfect competition, a profit maximizing firm chooses employment to satisfy the factor market equilibrium[17]

$$w = \frac{1}{\mu} Y_H = \frac{\nu}{\mu} A \left(\frac{H}{N}\right)^{\nu-1} = \frac{\nu}{\mu} \frac{y+\phi}{h} \tag{21}$$

Where $\mu \equiv \frac{\theta}{\theta-1} \in [1, \infty)$ is the markup, which is 1 with perfect competition when products are perfectly substitutable $\theta \to \infty$, so demand curves are perfectly elastic.[18] The labor demand curve will be increasing, decreasing or constant depending on the shape of the MPL schedule

$$w_H = \frac{1}{\mu} Y_{HH} = (\nu - 1)\frac{w}{H} \gtreqless 0 \iff \nu \gtreqless 1 \tag{22}$$

We assume the degree of increasing returns to labor is bounded above by the degree of monopoly power. This ensures the second-order condition for profit maximization holds.

Lemma 1. *$\nu < \mu$ is a sufficient condition for the second-order profit maximization condition to hold.*

Later we show it is necessary and sufficient for steady-state existence.[19] This restriction implies that for profit maximizing output MR must intersect MC from above (the second order condition for profit maximization). A higher degree of monopoly μ (more differentiated products) implies steeper

[17] The result follows from the profit maximization problem outlined in Appendix A.4. In the increasing returns case $\nu > 1$, the second-order condition for profit maximization is not always satisfied, so we give a necessary condition for this. However, our later condition $\nu < \mu$ is sufficient for this second-order necessary condition to hold.

[18] Labor demand h will vary depending on returns to scale. The relationship captures *'aggregate labor demand'* (Jaimovich 2007), the right-hand side is the marginal revenue product of labor which is the inverse of the markup multiplied by the marginal product of labor. The number of firms affects the relationship through the marginal product of labor since the markup is fixed. With endogenous markups and constant returns to scale, the number of firms also affect the MRPL (also true of LOV). Both can create upward sloping marginal product schedule $dw/dH > 0$.

[19] Hornstein 1993; Devereux, Head, and Lapham 1996 provide similar conditions in instantaneous-entry, zero-profit models with returns to scale.

MR which allows steeper downward sloping MC (higher ν). Horizontal MC only exists if MR is downward sloping, so some monopoly power exists. Increasing marginal costs $\nu < 1$ is compatible with any level of imperfect competition $\mu \in [1, \infty)$ including perfect competition.

Operating profits and output per firm (thus labor per firm) are isomorphic since $\pi = y - \frac{\nu}{\mu}(y + \phi)$ hence

$$\pi = y\left(1 - \frac{\nu}{\mu}\right) - \frac{\nu}{\mu}\phi = Ah^{\nu}\left(1 - \frac{\nu}{\mu}\right) - \phi \tag{23}$$

$$y = \frac{\mu\pi + \nu\phi}{\mu - \nu} \tag{24}$$

$$h = \left(\frac{y + \phi}{A}\right)^{\frac{1}{\nu}} = \left(\frac{\mu(\pi + \phi)}{A\,(\mu - \nu)}\right)^{\frac{1}{\nu}} \tag{25}$$

Operating profits respond proportionally but strictly less than output $0 < \pi_y = 1 - \frac{\nu}{\mu} < 1$. The implication is that economic profits are less volatile than output, and lemma 1 implies that this relationship cannot be negative.

2.3 Labor Market Equilibrium

In labor market equilibrium labor supply (8) equals labor demand (21): $H^{\eta}\bar{C} = \frac{A\nu}{\mu}H^{\nu-1}N^{1-\nu}$.[20] It is useful to write as a function of (N, λ)[21]

$$H(\lambda, N) = \left(N^{1-\nu}\lambda\frac{\nu A}{\mu}\right)^{\frac{1}{1+\eta-\nu}}, \quad 1 + \eta - \nu > 0 \tag{27}$$

[20]If labor is indivisible ($\eta = 0$) then all wage is consumed $\bar{C} = \frac{1}{\mu}A\nu h^{\nu-1}$, so there is no substitution effect. With constant marginal costs $\nu = 1$ then $\bar{C} = A/(\mu H^{\eta})$ there is only an income effect as wage is fixed. Jaimovich 2007 studies the effect of instantaneous entry on this relationship with both constant returns and indivisible labor, but N affects the relationship through endogenous markups $\mu(N)$ which causes indeterminacy.

[21]If we substitute out $N = H/h = H\left(\frac{A(\mu-\nu)}{\mu(\pi+\phi)}\right)^{\frac{1}{\nu}}$ in (27) we get labor as a function of profits

$$H = \left[\left(\frac{A(\mu - \nu)}{\mu(\pi + \phi)}\right)^{\frac{1-\nu}{\nu}}\frac{\lambda\nu A}{\mu}\right]^{\frac{1}{\eta}} \tag{26}$$

Whether labor increases, decreases or does not respond to a change in profits depends on returns to scale ν.

Lemma 2 (Labor Market Equilibrium Existence)**.** *To ensure that the labor market condition is well-defined* $\nu < 1 + \eta$

The restriction $\nu < 1 + \eta$ implies that the slope of the labor supply curve exceeds the slope of the labor demand curve. The labor supply curve slope is $\frac{dw}{dH} = \frac{\eta w}{H}$, and upward sloping in (H, w) space (or flat with indivisible labor $\eta = 0$). This must be greater than the slope of labor demand (marginal (revenue) product schedule) which is $\frac{dw}{dH} = w_h h_H = \frac{(\nu-1)w}{h}\frac{1}{N} = \frac{(\nu-1)w}{H}$. As noted, demand for labor can be upward sloping if returns to labor are increasing $\nu > 1$.

Proposition 1 (Existence)**.** *Necessary and sufficient condition for existence*

$$\nu < \min[\mu, 1 + \eta] \tag{28}$$

A sufficient condition is that there are increasing marginal costs $\nu < 1$*. Where* $1 + \eta > 1$ *because we rule out indivisible labor* $\eta = 0$*.*

Proof. Combine profit existence Lemma 1 and labor market existence Lemma 2. □

Entry alters employment per firm which, through marginal costs, affects the efficiency of labor and thus the real wage it is paid. With a decreasing MPL, entry increases the real wage and hence labor supply; with increasing MPL the opposite holds.

Proposition 2 (General Equilibrium Labor Behavior)**.** *From the labor market equilibrium condition* (27)*, we can see that labor responses to entry are*

$$H_N > 0 \iff \nu \in (0, 1) \tag{29}$$
$$H_N = 0 \iff \nu = 1 \tag{30}$$
$$H_N < 0 \iff \nu \in (1, \infty) \tag{31}$$

In deriving this result we show that labor elasticity to number of firms

$\varepsilon \equiv H_N \frac{N}{H}$ is constant and bounded

$$\varepsilon = \frac{1 - \nu}{1 + \eta - \nu} \tag{32}$$

It is bounded by $\frac{-\eta}{1-\nu+\eta} < \varepsilon < 1$. The upper bound occurs with indivisible labor $\eta \to 0$. The lower bound follows from $\nu < 1 + \eta$ so that (working right to left) $\frac{-\eta}{1-\nu+\eta} < \frac{1-(1+\eta)}{1-\nu+\eta} < \frac{1-\nu}{1-\nu+\eta} = \varepsilon$. If $\nu = 1$ then $\varepsilon = 0$. If $\nu < 1$ then $0 < \varepsilon < 1$. And if $\nu > 1$ then $-\infty < \varepsilon < 0$.[22]

2.3.1 Total Derivatives: Labor Effect Vs. Business Stealing

In section 2.1 we derived the partial derivatives of aggregate output with respect to labor $Y_H > 0$ and firms $Y_N \gtreqless 0$, assuming H and N were independent. Now that we have determined $H(\lambda, N)$ we can assess total derivatives of output with respect to entry by considering that labor changes endogenously. Understanding this mechanism is important for our results on the effect of entry on aggregate output to be derived later. The main point is that entry has an ambiguous effect on aggregate output if there are decreasing returns $\nu < 1$ so that $\varepsilon > 0$. This is because entry strengthens labor supply which can increase output. Whereas with constant or increasing returns $\nu \geq 1$ an entrant always decreases aggregate output.

$$\frac{dY}{dN} = y + N\frac{dy}{dN} = \varepsilon(1 + \eta)Ah^{\nu} - \phi \tag{33}$$

The first equality states that an entrant contributes its own output y but has a *business stealing* (Mankiw and Whinston 1986) effect on the output of all other incumbents. In the appendix we show this business stealing effect is strictly negative $N\frac{dy}{dN} = \nu(y + \phi)(\varepsilon - 1) < 0$. The second equality of (33) emphasizes the role of firm level returns to scale. It states that an entrant has a negative effect by bringing in an extra fixed cost, but it has another positive, negative or zero effect depending on the labor elasticity to entry ε.

The aggregate flow of operating profits given w equals $N\pi$, where π is

[22] See Appendix A.5 for proof.

firm level profit.[23]

$$\pi = Y_N + \left(1 - \frac{1}{\mu}\right) Y_H \frac{H}{N} \tag{34}$$

In terms of profits this can be written $\frac{dY}{dN} = Y_N + Y_H H_N = \pi - \left(1 - \frac{1}{\mu} - \varepsilon\right) Y_H h$ which is useful when we analyze zero-profit steady state.[24] The first term is the partial derivative effect of an entrant (19) which we have explained is ambiguous based on ν, and the second term is the labor response which is also ambiguous based on ν. Since y and π are in a one-one relationship, the business stealing effect can also be interpreted as entrants diminishing profits, from (23) $\frac{d\pi}{dN} = \frac{dy}{dN}\left(1 - \frac{\nu}{\mu}\right) < 0$. In the dynamic analysis we shall use the expression for dividends with $H(\lambda, N)$ substituted out:

$$\pi(\lambda, N) = \left(\frac{A^{1+\eta}(\nu\lambda)^{\nu}}{\mu^{1+\eta} N^{\eta\nu}}\right)^{\frac{1}{1+\eta-\nu}} (\mu - \nu) - \phi \tag{35}$$

2.4 The Entry Decision

What determines the number of firms operating at each instant t? We develop a congestion effects model of firm entry such that at time t there is a flow cost of entry $q(t)$ which increases in net entry $E(t)$.[25]

$$E(t) \equiv \dot{N} \tag{36}$$

$$q(t) = \gamma E(t) \tag{37}$$

The sensitivity to congestion parameter $\gamma \in (0, \infty)$ represents red tape or regulation in firm creation. Filing papers or gaining accreditation makes start-ups more sensitive to flows of entry as regulator's offices become more

[23] The result follows from substituting w (21) and Y (18) out of the aggregate profit expression $N\pi = Y - wH$ such that $N\pi = Y_N N + Y_H H - \frac{Y_H}{\mu} H$, which rearranges to (34).

[24] See Appendix for full derivation and discussion.

[25] Entry and exit are symmetric, with $-q$ being the cost of exit at time t. There are sunk costs to entry and dismantling fees, such as severance payments, to exit. See Das and Das 1997; Datta and Dixon 2002 for further details. Exit and entry symmetry is not essential, exit could require a fixed cost, perhaps zero, as in Das and Das 1997 and Hopenhayn 1992 or evolve endogenously according to productivity Melitz 2003; Hamano and Zanetti 2017.

congested (i.e. a queuing cost). Aggregating across all entry in a period gives a quadratic firm entry adjustment cost function

$$\mathcal{C}(E) \equiv \int_0^E q \, dE = \frac{\gamma}{2} E^2 = \frac{q^2}{2\gamma} \tag{38}$$

$\mathcal{C}(E)$ is a non-negative, convex function of the rate of entry. With zero entry, the aggregate cost and marginal cost of firm creation is zero $\mathcal{C}(0) = \mathcal{C}_E(0) = 0$. The interpretation of modelling the aggregate sunk cost as an adjustment cost is that firm creation and destruction, whether positive (net entry) or negative (net exit), generates resource costs.

The flow of entry in each instant is determined by an *arbitrage condition* that equates the return on bonds (opportunity cost of entry) with the return on setting up a new firm. It is a differential equation in q, which determines the entry flow by (37).[26]

$$\frac{\pi}{q} + \frac{\dot{q}}{q} = r \tag{39}$$

π is given by (34) which will make this a nonlinear differential equation in N.[27] The first left-hand side term is the number of firms per dollar $(1/q)$ times the flow operating profits (dividends) the firm will make if it sets up. The second term reflects the change in the cost of entry. If $\dot{q}/q > 0$, then it means that the cost of entry is increasing, so that there is a capital gain associated with entry at time t if $\dot{q}/q < 0$ it means entry is becoming cheaper, thus discouraging immediate entry. The sunk cost $q(t)$ represents the net present value of incumbency: it is the present value of profits earned if you are an incumbent at time t.[28] This arises since the entrants are indifferent between entering and staying out. When $q < 0$, the present value of profits is negative: in equilibrium this is equal to the cost of exit. In steady state, we

[26]The arbitrage equation can be written in a way directly analogous to the user cost of capital $\pi = q\left(r - \frac{\dot{q}}{q}\right)$ in capital adjustment cost models.

[27]Note that our entry model has the standard models as limiting cases: when $\gamma = 0$, we have instantaneous free entry so that (39) becomes $\pi = 0$ and there are zero profits each instant. If $\gamma \to +\infty$, then changes in N become very costly and N moves little if at all which approximates the case of a fixed number of firms.

[28]This is because of the free-entry assumption that sunk costs equal the net present value of the firm. See Stokey 2008 for a general discussion.

have $E = q = 0$, so that the entry model implies the zero-profit condition. Entry costs only arise on convergence to steady state.

Accounting for entry costs, aggregate profits Π are the operating profits (dividends) of firms less the entry costs paid by the entrants

$$\Pi = N\pi - \gamma\frac{E^2}{2} = NY_N + \left(1 - \frac{1}{\mu}\right) Y_H H - \gamma\frac{E^2}{2} = Y(N, H) - wH - \frac{q^2}{2\gamma} \quad (40)$$

3 Equilibrium, Steady State and Solution

The economic system is five dimensional $\{\lambda, N, q, B, H\}$ with four differential equations and one static equation. The static intratemporal condition (27) implies $H(\lambda, N)$, so the system can be reduced to four differential equations in four unknowns, and since the consumption differential equation implies consumption is constant $\lambda(t) = \bar{\lambda}$, we have three dynamic equations in N, q, B.

$$\dot{\lambda} = 0 \implies \lambda(t) = \bar{\lambda}$$

$$\dot{N}(q) = \frac{q}{\gamma} \quad (41a)$$

$$\dot{q}(N, \bar{\lambda}, q) = rq - \pi(N, H(\bar{\lambda}, N)) \quad (41b)$$

$$\begin{aligned} \dot{B}(B, N, \bar{\lambda}, q) &= rB + wH(\bar{\lambda}, N) + \Pi(N, H(\bar{\lambda}, N), q) - \bar{C}(\bar{\lambda}) - G \\ &= rB + Y(N, H(\bar{\lambda}, N)) - \mathcal{C}(q) - \bar{C}(\bar{\lambda}) - G \end{aligned} \quad (41c)$$

Accompanying the differential equations in system (41) there are three boundary conditions: the household transversality (9); the initial condition on bonds; the initial condition on number of firms. Notably the industry dynamics (N, q) form a two dimensional subsystem of the three dimensional system, with bonds being B determined through (41c) alone. Therefore we shall solve recursively: first solving the industry dynamics subsystem for $N(t), q(t)$, then solve for bonds $B(t)$ based on these solutions.

3.1 Steady-state

Steady state is non-standard because there are three steady state conditions $\dot{N} = \dot{q} = \dot{B} = 0$ but four unknowns $\bar{\lambda}, q, N, B$.[29] In order to get an extra equation to solve this system for steady state, first we find a solution to the dynamic system for its timepaths of $N(t, \bar{\lambda}), q(t, \bar{\lambda}), B(t, \bar{\lambda})$ conditional on knowing one steady-state variable $\bar{\lambda}$. Second we use the limit of the bond solution and transversality to acquire an extra steady state condition, allowing us to solve for steady state. It is this procedure which causes steady state to depend on initial conditions N_0, B_0, so-called path dependency or hysteresis.[30]

We use a tilde to denote a steady state variable. The $\dot{N} = 0$ differential equation immediately implies that steady-state sunk costs are zero, which equivalently implies the net present value of a firm in steady state is zero.

$$\tilde{q} = 0 \tag{42}$$

This leaves two steady-state conditions $\dot{q} = \dot{B} = 0$ in three unknowns $\tilde{N}, \bar{\lambda}, \tilde{B}$. Through the arbitrage condition (41b), zero sunk costs (42) imply operating profits are zero

$$\tilde{\pi} = 0 \tag{43}$$

The zero profit condition determines labor per firm (or aggregate labor as a linear function of number of firms $\tilde{H}(\tilde{N})$)

$$\tilde{h} = \left(\frac{\mu \phi}{A(\mu - \nu)} \right)^{\frac{1}{\nu}} \tag{44}$$

[29] This occurs because the consumption differential equations is always in steady-state ($\dot{\lambda} = 0$) due to perfect consumption smoothing from $r = \rho$ which implies consumption is fixed $\lambda = \bar{\lambda}$, but it does not relate to other variables in the system.

[30] An implication of this feature is that temporary shocks may have permanent effects.

Labor per firm determines output per firm and wage[31]

$$\tilde{y} = \frac{\nu}{\mu - \nu}\phi \tag{45}$$

$$\tilde{w} = \left(\frac{A}{\mu}\right)^{\frac{1}{\nu}} \nu \left(\frac{\phi}{\mu - \nu}\right)^{1-\frac{1}{\nu}} \tag{46}$$

With $\tilde{h}$ and $\tilde{w}$ determined by the free entry arbitrage condition $\tilde{\pi} = 0$, then the labor market equilibrium condition (27) determines the number of firms as a function of the consumption index, and therefore labor as a function of consumption index:

$$\tilde{N}(\bar{\lambda}) = \frac{(\bar{\lambda}\tilde{w})^{\frac{1}{\eta}}}{\tilde{h}} \tag{47}$$

$$\tilde{H}(\bar{\lambda}) = (\bar{\lambda}\tilde{w})^{\frac{1}{\eta}} \tag{48}$$

In order to find $\bar{\lambda}$, we are left with one steady-state condition $\dot{B} = 0$ that we have not used: the output market clearing condition (steady-state bond accumulation equation).

$$G + \bar{C}(\bar{\lambda}) - \tilde{w}\tilde{H}(\bar{\lambda}) - r\tilde{B} = 0 \tag{49}$$

This is an excess demand function for the steady state in terms of the price of marginal utility $\bar{\lambda}$. The first two terms $G + C(\bar{\lambda})$ represent expenditure and are decreasing in $\bar{\lambda}$. The second two terms $wH(\bar{\lambda}) + r\tilde{B}$ represent income and are increasing in $\bar{\lambda}$. By the intermediate value theorem, this implies that there exists a $\bar{\lambda} > 0$ such that the economy is at the steady state equilibrium given $\tilde{B}$ (See Appendix A.8 for proof of existence and uniqueness with endogenous $\tilde{B}(\bar{\lambda})$.).

In this section we partly defined steady-state $\{\tilde{N}, \bar{\lambda}, \tilde{B}\}$ for the primitive variables of the dynamical system $N, \bar{\lambda}, B$, given steady-state bonds $\tilde{B}$. We gave $\tilde{N}(\bar{\lambda})$ analytically in (47), then used (49) to prove a steady-state $\bar{\lambda}$ must exist given $\tilde{B}$. In the next section, we derive solutions for dynamics which

[31] Since zero profits imply $0 = \tilde{y} - \tilde{w}\tilde{h}$ then steady-state wage is equivalent to labor productivity $\tilde{w} = \frac{\tilde{y}}{\tilde{h}}$.

provide an additional steady-state condition $\tilde{B}(\bar{\lambda})$ that teamed with (49) and (47) can solve for $\bar{\lambda}$ by expressing (49) entirely in $\bar{\lambda}$ terms

$$G + \frac{1}{\bar{\lambda}} - \tilde{w}^{1+\frac{1}{\eta}}\bar{\lambda} - r\tilde{B}(\bar{\lambda}) = 0$$

3.2 Linearized system

The analysis of the steady state was conditional on the level of steady state bonds $\tilde{B}$. However to determine $\tilde{B}$ we need to know the path taken to equilibrium. The dynamics of the system will be analyzed by linearizing around the steady state. Where the 3×3 matrix is the Jacobian $\mathbf{J}$, the linearized system is[32]

$$\begin{bmatrix} \dot{N} \\ \dot{q} \\ \dot{B} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{\gamma} & 0 \\ \frac{1}{\tilde{N}(\bar{\lambda})}\frac{\nu\eta\phi}{1+\eta-\nu} & r & 0 \\ \tilde{\Omega} & 0 & r \end{bmatrix} \begin{bmatrix} N(t) - \tilde{N} \\ q(t) - \tilde{q} \\ B(t) - \tilde{B} \end{bmatrix} \tag{50}$$

$$\text{where} \quad \tilde{\Omega} = \frac{\nu\phi\mu}{\mu - \nu}\left(\varepsilon - 1 + \frac{1}{\mu}\right) \tag{51}$$

Since the total effect of an entrant on aggregate output is an important mechanism for our analysis we denote it

$$\Omega \equiv \frac{dY}{dN}$$

The ambiguous effect of entry on aggregate output ($\Omega \gtreqless 0$) explored away from steady state in section 2.3.1 is also ambiguous in steady state ($\tilde{\Omega} \gtreqless 0$). It depends on $\left(\varepsilon - 1 + \frac{1}{\mu}\right)$. We discuss this extensively in section 3.2.3. For dynamics it implies that the Jacobian element corresponding to the effect of entry on bond accumulation $\frac{d\dot{B}}{dN}|^{\tilde{}} = \tilde{\Omega}$ is ambiguous.

[32]Detailed derivation in Appendix A.6

3.2.1 Number of Firms and Entry (industry dynamics) Solution

The determinant and trace of the industry dynamics $\{N, q\}$ sub-system $\mathbf{B} \in \mathbb{R}^2$ in (50) are

$$\det(\mathbf{B}) = \Delta = \frac{\frac{d\tilde{\pi}}{dN}}{\gamma} = -\frac{\nu\eta\phi}{\gamma\,(1+\eta-\nu)\,\tilde{N}(\bar{\lambda})} < 0 \tag{52}$$

$$tr(\mathbf{B}) = r \tag{53}$$

$\det(\mathbf{B})$ is negative as $1+\eta > \nu$ and is increasing in $\bar{\lambda}$.[33] The root to the characteristic polynomial corresponding to the subsystem is

$$\Gamma(\bar{\lambda}) = \frac{r}{2}\left(1 \pm \frac{1}{r}\left[r^2 - 4\Delta(\tilde{N}(\bar{\lambda}))\right]^{\frac{1}{2}}\right) \tag{54}$$

The discriminant (square root term) is positive since the determinant is negative ($\Delta < 0$). This implies two distinct real roots. And since the discriminant exceeds 1, then so does its square root so there will be one positive and one negative root. Hence the system is saddle-path stable, with a negative real root Γ and a positive real root Γ^U. Furthermore the trace is positive so the sum of the eigenvalues is positive implying the positive eigenvalue is larger than the absolute value of the negative eigenvalue. Our focus is the stable root which is negative

$$\Gamma = \frac{1}{2}\left(r - \left[r^2 - 4\Delta\right]^{\frac{1}{2}}\right) \tag{55}$$

Lemma 3. *The stable eigenvalue is increasing in* $\bar{\lambda}$

Proof. See Appendix A.7. □

The solution to the linearized subsytem is

$$N(t) = \tilde{N} + \exp[\Gamma(\bar{\lambda})t](N_0 - \tilde{N}) \tag{56}$$

take derivative to get the net entry rate $E = \dot{N} = \Gamma\exp[\Gamma t](N_0 - \tilde{N})$ and

[33] See Appendix A.7 for proof.

substitute $q = \gamma E$ for the sunk cost solution

$$q(t) = \gamma\Gamma \exp[\Gamma t](N_0 - \tilde{N}) \tag{57}$$

The derivative of the solution is $\dot{q} = \Gamma^2 \gamma \exp(\Gamma t)(N_0 - \tilde{N})$, so the growth (shrinkage) in the cost of entry (firm NPV) is given in absolute terms by the stable eigenvalue

$$\left|\frac{\dot{q}}{q}\right| = \Gamma$$

with the sign being determined by whether profits are positive (firms accumulation) or negative (decumulation).

3.2.2 Bonds Solution

Combining (41c) and (9) provides a condition that the solution for bonds must satisfy in the long run (full derivation Appendix A.3).

$$0 = B_0 + \int_0^\infty e^{-rt} \left[Y - \frac{q^2}{2\gamma} - C - G\right] dt \tag{58}$$

The two terms must cancel out, which has an intuitive interpretation. The first term is the initial position of bond holdings. $B_0 > 0$ implies the country begins as a borrower, $B_0 < 0$ implies it begins as a creditor. The second term represents trade surplus if positive and deficit if negative. Therefore (58) states that if a country begins as a borrower, at some point over the time horizon it must run a trade deficit.

Linearizing the differential equation in bonds gives

$$\dot{B}(t) = \tilde{\Omega}\left[N(t) - \tilde{N}\right] - \frac{\tilde{q}}{\gamma}\left[q(t) - \tilde{q}\right] + r\left[B(t) - \tilde{B}\right] \tag{59}$$

where $\tilde{q} = 0$. Then substitute in the $N(\bar{\lambda}, t)$ solution (56) restricts the differential equation to be a linear first-order nonhomogeneous differential equation in $B(t)$

$$\dot{B}(t) = \tilde{\Omega}\left[\exp[\Gamma t](N_0 - \tilde{N})\right] + r\left[B(t) - \tilde{B}\right] \tag{60}$$

If the economy starts with bonds $B(0) = B_0$ the solution to (60) is

$$B(t) = \tilde{B} + \frac{\tilde{\Omega}}{\Gamma(\bar{\lambda}) - r} \exp[\Gamma(\bar{\lambda})t](N_0 - \tilde{N}) \tag{61}$$

where $\frac{d\dot{B}}{dN}|^{\sim} = \tilde{\Omega}$ implies the effect of entry on aggregate output equals the effect of entry on the flow of bonds evaluated at steady state. $\tilde{\Omega}$ affects how accumulation of firms $N_0 \to \tilde{N}$ so $N_0 - \tilde{N} < 0$ changes stock of bonds $B(t)$. $\tilde{\Omega} > 0$ then entry strengthens home production and increases bond investment, whereas $\tilde{\Omega} < 0$ then entry weakens home production and decreases bond investment. In the Walrasian case $(\mu = 1, \nu < 1)$, $\tilde{\Omega} > 0$ and the accumulation of firms leads to a reduction in bonds. The main mechanism here is that there is a positive effect of N on labor supply and output $(Y_{HN} > 0)$, so that having too few firms means that wages, labor income and home production are below their steady state level. To maintain consumption, this low level of income is compensated by higher than steady state imports, financed by running down bonds. An *increase* in firms per se makes wages higher. However, the number of firms is increasing because it is below the steady-state. The stock of bonds decreases because entry implies that the initial level of N was low in the first place, not because the accumulation of firms lowers income.

However, given $\mu > 1, \nu < 1$, if μ is large enough then bonds will increase as firms are accumulated. This is because the level of profits along the path to equilibrium is large: whilst the number of firms is below equilibrium, the extra profits generated are enough to exceed the adjustment costs and lower wage. In addition, there is a capacity effect, so that productivity is higher whilst the number of firms is below equilibrium (for $\mu > 1$, free-entry leads to excessive number of firms in steady-state). In the case of $\nu \geq 1$, the flow of entry leads to an increase in the stock of bonds: this is because N has a negative effect on wages and profits, so that N below its steady state implies income above the steady state.

3.2.3 Effect of Entry on Aggregate Output

In steady state entry may increase, decrease or have no effect on aggregate output $\tilde{\Omega} \gtreqless 0$. This corresponds to whether entry increases, decreases or has no effect on labor supply, which depends on whether labor is employed with decreasing, increasing or constant returns.

Proposition 3 (Entry and Aggregate Output)**.** *The effect of entry on aggregate output $\tilde{\Omega}$ is ambiguous in steady-state.*

1. *Lack of Entry:* $\tilde{\Omega} > 0 \iff 1 - \nu > \eta(\mu - 1)$
2. *Excess Entry:* $\tilde{\Omega} < 0 \iff 1 - \nu < \eta(\mu - 1)$
3. *Optimal Entry:* $\tilde{\Omega} = 0 \iff 1 - \nu = \eta(\mu - 1)$

For $\nu \geq 1$ there is always excessive entry $\tilde{\Omega} < 0$. For $\nu < 1$ all outcomes are possible.[34]

Next we provide a discussion of the three possible cases.[35] From the proof the outcome depends on whether the negative business stealing effect $-\left(\frac{\mu-1}{\mu}\right) \leq 0$, $\mu \in [1, \infty)$ dominates the labor elasticity to entry effect $\frac{-\eta}{1+\eta-\nu} < \varepsilon < 1$, which may be positive, negative or zero.

Excess Entry $\tilde{\Omega} < 0$: If there are constant $\nu = 1$ or increasing $\nu > 1$ returns to labor, $\varepsilon \leq 0$, then the fall in labor reinforces the negative business stealing effect, so there is unambiguously a negative effect of entrants on aggregate output in steady state. This is a sufficient condition but is not necessary, providing the business stealing effect is large enough it can override even a positive labor elasticity effect that arises with decreasing returns $\nu < 1$.

[34] Optimal entry refers to the number of firms that maximizes steady-state aggregate output, conditional on a markup existing. There is no maximum with perfect competition $\mu = 1$, always a lack of entry due to a positive labor effect and no negative markup (business stealing) effect.

[35] Etro 2009; Etro and Colciago 2010 provide a discussion of 'golden rule' number of firms when there is endogenous imperfect competition, constant returns and love-of-variety. The golden rule number of firms is that which maximizes consumption and therefore output in steady-state. They show that imperfect competition causes excessive entry in steady-state, which our proposition corroborates ($\mu > 1$ and $\nu = 1$ implies $1 - \nu < \eta(\mu - 1)$, so excess entry).

1. Example: Positive labor elasticity effect, dominated by negative business stealing effect $\nu = 0.9,\ \eta = 1$ therefore $\varepsilon = 0.\overline{09}$ with $\mu = 1.15$ business stealing is -0.13.

2. Constant Returns Special Case $\nu = 1$: The labor effect is zero, so only the negative business stealing effect is present. The smaller the markup $\mu \to 1$ the smaller the negative business stealing effect. But it cannot equal 1 due to the existence condition $\nu < \mu$.

With large markups this outcome is likely. With less divisible labor $\eta \to 0$ this outcome is more likely.

Lack of Entry $\tilde{\Omega} > 0$: If there are decreasing returns $\nu < 1$ then $0 < \varepsilon < 1$ and the boost in labor from entry works against the negative business stealing effect, so there can be too little entry if this positive effect dominates the negative business stealing effect. $\varepsilon > 0$, hence $\nu < 1$, is necessary but not sufficient, sufficiency requires it is positive *and* larger than the negative business stealing effect.

1. Example: Positive labor elasticity effect dominates negative business stealing effect $\nu = 0.9,\ \eta = 1$ therefore $\varepsilon = 0.\overline{09}$ with $\mu = 1.05$ business stealing is -0.05.

2. Perfect Competition Special Case $\mu = 1, \nu < 1, \tilde{\Omega} > 0$: There is no negative business stealing effect, and the the existence condition $\nu < \mu$ enforces decreasing returns. Therefore entry always has a positive effect, implying lack of entry in steady state in the Walrasian (perfect competition) economy.

Optimal Entry $\tilde{\Omega} = 0$: A necessary condition is that the ambiguous labor elasticity effect is positive $\varepsilon > 0$, so it can counterbalance the negative business stealing effect. Therefore a necessary condition is decreasing returns $\nu < 1$.

1. Example: $\nu = 0.9,\ \eta = 1,\ \mu = 1.1$

3.3 Steady-state Bonds

The linearized dynamics give an explicit solution for steady state bonds as a function of $\bar{\lambda}$ and the initial conditions N_0, B_0. Evaluate (61) at $t = 0$ implies

$$\tilde{B}(\bar{\lambda}) = B_0 - \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}(N_0 - \tilde{N}(\bar{\lambda})) \tag{62}$$

therefore the steady-state bond condition (62) and steady-state arbitrage condition (47) give the excess demand condition (49) in terms of $\bar{\lambda}$ only

$$\tilde{w}\tilde{H}(\bar{\lambda}) + r\tilde{B}(\bar{\lambda}) - \bar{C}(\bar{\lambda}) - G = 0 \tag{63}$$

We can solve this for the steady-state consumption index $\bar{\lambda}$, which then provides $\tilde{C}(\bar{\lambda}), \tilde{H}(\bar{\lambda}), \tilde{N}(\bar{\lambda})$, $\tilde{B}(\bar{\lambda})$. We cannot solve (63) analytically since it is highly nonlinear in $\bar{\lambda}$. However we can show analytically that a unique solution exists, and then solve for this numerically. A useful lemma to show uniqueness (and other results) is that the steady-state excess demand function is strictly increasing in inverse consumption, so is decreasing in consumption given N_0 begins within a neighbourhood of $\tilde{N}$.

Lemma 4 (Excess Demand Monotonically Increasing)**.** *The steady-state market-clearing condition is monotonically increasing in* $\bar{\lambda}$

$$\tilde{w}\frac{d\tilde{H}}{d\lambda} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\tilde{C}}{d\bar{\lambda}} > 0 \tag{64}$$

if the following sufficient condition holds

$$\left(\varepsilon - 1 + \frac{1}{\mu}\right)\left(\frac{N_0}{\tilde{N}(\bar{\lambda})} - 1\right) \geq -\left(\frac{\varepsilon - 1}{\Gamma(\bar{\lambda})} + \frac{1}{r\mu}\right)(r - 2\Gamma(\bar{\lambda})) \tag{65}$$

Proof. See appendix A.8. □

The right-hand side of (65) is strictly negative and the left-hand side is ambiguous. This condition is weaker than the simpler sufficient condition $N_0 - \tilde{N}(\bar{\lambda}) \to 0$ which is commonly assumed and ensures the left-hand side is zero.[36] The condition always holds if there is entry $N_0 < \tilde{N}$ and $\varepsilon - 1 + \frac{1}{\mu} < 0$ (i.e. $\tilde{\Omega} < 0$) implying the left-hand side is positive.

Corollary 1 ($\bar{\lambda}$ Uniqueness)**.** *If* (65) *holds then there is a unique* $\bar{\lambda}$ *that solves* (64)*.*

Proof. Lemma 4 shows that given (65) the steady state market clearing condition (excess of income over expenditure) is strictly monotonic in $\bar{\lambda}$. Hence, if a steady-state exists it is a *unique* steady state solution for $\bar{\lambda}$. □

4 Technological Change

4.1 Comparative Statics

An improvement in technology A reduces employment per firm but output per firm (firm scale) (12) is unaffected. Consequently an improvement in technology increases wages[37]

$$\frac{d\tilde{h}}{dA} = -\frac{\tilde{h}}{\nu A} < 0 \tag{66}$$

$$\frac{d\tilde{w}}{dA} = \frac{\tilde{w}}{\nu A} > 0 \tag{67}$$

Therefore in the long run technological progress crowds-out labor at the product-level but output is unaffected (aggregate output will expand as there are more products each requiring less labor). These comparative statics are simple as they only depend on exogenous variables. However, the aggregate endogenous variables $\{\bar{C}, \tilde{N}, \tilde{B}\}$ ((7), (47), (62)), excluding $\tilde{q}$ which is zero, are a function of A directly but also depend on $\bar{\lambda}(A)$. Therefore technology change has a direct (partial) and an indirect (consumption) effect.[38]

[36] See Turnovsky 1997, p.68 (footnote 8) for a justification of this.

[37] An increase in steady-state wages is equivalent to an increase in labor productivity since $\tilde{w} = \frac{\tilde{y}}{\tilde{h}}$.

[38] We call the indirect effect a consumption effect as $\bar{\lambda}(A)$ is inverse consumption by (7).

Proposition 4 (Long-run Effect of Technology)**.** *A permanent improvement in technology:*

$$\frac{d\bar{C}}{dA} > 0 \tag{68}$$

$$\frac{d\tilde{N}}{dA} > 0 \tag{69}$$

$$\operatorname{sgn} \frac{d\tilde{B}}{dA} = \operatorname{sgn} -\tilde{\Omega} \tag{70}$$

$$\operatorname{sgn} \frac{d\tilde{H}}{dA} = \operatorname{sgn} \left[B_0 - \frac{\tilde{\Omega}}{\Gamma - r} N_0 \right] \tag{71}$$

$$\frac{d\tilde{Y}}{dA} = \tilde{y}\frac{d\tilde{N}}{dA} > 0 \tag{72}$$

From the steady-state market clearing condition, the implicit function theorem implies that technology unambiguously increases consumption. This rise in consumption (indirect effect) decreases aggregate labor and number of firms, whereas the direct partial effects of increased technology increase labor and number of firms. Overall, the partial effect dominates in the number of firms case, whereas it is ambiguous in the labor case. The increase in the stock of firms implies an increase in aggregate output, and a bond response that depends on the whether there are excessive, insufficient or optimal number of firms. The effect on the labor supply is ambiguous because there is a conflict of income and substitution effects: the higher wage causes a substitution effect for less leisure and more consumption, which increases labor. Whereas the income effect increases leisure and decreases labor. Which effect dominates depends on the level of initial wealth. From (62) $B_0 - \frac{\tilde{\Omega}}{\Gamma - r} N_0$ is the initial value of wealth in terms of bonds.[39] If $\tilde{\Omega} > 0$, that is $\nu < 1$ and μ small enough, then a sufficient condition for employment to increase $\frac{d\tilde{H}}{dA} > 0$ is that bond holdings are non-negative $B_0 \geq 0$. Likewise, if $\tilde{\Omega} < 0$, (for which $\nu \geq 1$ is sufficient) then a sufficient condition for employment to decrease $\frac{d\tilde{H}}{dA} < 0$ is that bond holdings are non-positive $B_0 \leq 0$.

[39] From (62), $-\frac{\tilde{\Omega}}{\Gamma - r} N_0 = \tilde{B} - B_0 - \frac{\tilde{\Omega}}{\Gamma - r}\tilde{N}$ thus the term $-\frac{\tilde{\Omega}}{\Gamma - r} N_0$ is the present value of the bonds that would have been decumulated/accumulated if $\tilde{N} = 0$.

Bonds respond in the opposite direction to the entry effect on output. If technology-induced entry increases GDP, then bonds decrease (less borrowing is necessary). If technology-induced entry decreases GDP, then bonds increase (more borrowing is necessary). Since steady-state bonds only depend on technology through $\tilde{N}$, the bond response follows the number of firms increase:m $\frac{d\tilde{B}}{dA} = \frac{d\tilde{B}}{dN}\frac{d\tilde{N}}{dA}$, and to a first-order approximation sgn $\frac{d\tilde{B}}{dN} \approx$ sgn $-\tilde{\Omega}$.[40] Similarly the increase in number of firms determines that aggregate output increases as long-run output per firm (firm scale) is constant.

4.2 Comparative Dynamics

From the dynamic solution for number of firms (56), we can see that on impact $t = 0$ of a shock the number of firms is fixed $N(0) = N_0$, whereas entry adjusts $E(0) = \Gamma(N_0 - \tilde{N})$, which affects the stock of firms an instance later. In other words number of firms is a stock (state) variable, and entry is a flow (jump) variable. Thus entry jumps the economy onto its stable manifold instantaneously as the shock hits, subsequently the number of firms responds as the economy evolves along this manifold. Therefore the difference between the impact and long-run effects depend on the effect of entry.

Proposition 5. *On impact of a technology shock hours and wages will increase, decrease or remain constant relative to their long-run level depending on whether labor returns to scale are increasing $\nu > 1$, decreasing $\nu < 1$ or constant $\nu = 1$.*

$$\text{sgn}\left[\frac{dH(0)}{dA} - \frac{dH(\infty)}{dA}\right] = \text{sgn}\left[\nu - 1\right] \tag{73}$$

$$\text{sgn}\left[\frac{dw(0)}{dA} - \frac{dw(\infty)}{dA}\right] = \text{sgn}\left[\nu - 1\right] \tag{74}$$

On impact the labor effect is ambiguous, as in the long run, due to competing substitution and income effects. The reason is also the same (income and substitution effects may clash). However, if we look at the difference

[40] The approximation arises from assuming we begin close to steady-state $N_0 - \tilde{N} \to 0$. From (62) removes the effect of the eigenvalue responding to $\tilde{N}$.

between the impact and long-run effect, this depends on whether there is an increasing or decreasing MPL at the firm level. When $\nu < 1$, on impact there is a negative relationship between the real wage and employment.; when $\nu > 1$ a positive relation; when $\nu = 1$ no relation. We can thus get undershooting of employment ($\nu > 1$) or overshooting ($\nu < 1$) on impact relative to the new long-run level depending on whether entry increases or decreases the marginal product.

Table 1 captures the combination of static (Proposition 4) and dynamic effects (Proposition 5) on labor. Rows capture the static effect that labor might in the long-run increase, decrease or remain constant depending on initial wealth. Columns capture the dynamic effect that labor might initially overshoot, undershoot or perfectly reflect its long-run level.

	$\nu < 1$	$\nu > 1$	$\nu = 1$
$B_0 > \frac{\tilde{\Omega}}{\Gamma - r} N_0$	Increase, Overshoot	Increase, Undershoot	Increase, Constant
$B_0 < \frac{\tilde{\Omega}}{\Gamma - r} N_0$	Decrease, Overshoot	Decrease, Undershoot	Decrease, Constant
$B_0 = \frac{\tilde{\Omega}}{\Gamma - r} N_0$	Constant, Overshoot	Constant, Undershoot	Constant, Constant

Table 1: Conditions for Taxonomy of Labor Dynamics

4.3 Reconciling with Empirical Evidence

In the theoretical model we derived the result that the short-run response of labor depends on whether the marginal product of labor is increasing or decreasing. In most models of entry, such as Bilbiie, Ghironi, and Melitz 2012, there is a constant marginal product of labor, so that there is no short-run impact on labor. Chang and Hong 2006 conduct an SVAR analysis of labor responses to technology shocks across US manufacturing industries. They show that of their 2-digit industry estimates, 14 industries show a positive response (4 significant) while 6 industries show a negative response (1 significant).[41] Additionally they provide estimates of returns to scale using the

[41] *Instruments* and *Non-electornic* are zero at 3 decimal places but positive with greater precision. Statistical significance is at the 10% level. *Misc* are significant with greater precision than reported in Table 2: $\frac{SRR}{SD} = 0.01626/0.0098 = 1.6492 > t^{crit.} = 1.6449$.

methodology of Basu, Fernald, and Kimball 2006 (BFK). The BFK methodology is to run a log-linear regression of output on inputs with a common coefficient γ on capital and employment for each industry, with an additional coefficient β on hours per worker.[42] The coefficient γ is interpreted as returns to scale which is reported by Chang and Hong (Table 5) for their dataset. In terms of our model, in which there is only labor, we can interpret the increasing or decreasing marginal product of labor $\nu \gtreqless 1$ either as the coefficient γ (i.e. interpreting labor input as employment) or as the sum of the coefficients γ and β (i.e. the coefficient on total hours, the product of employment and hours-per-worker). Chang and Hong (Table 5) provide estimates of γ for 20 two-digit industries (ten durables and ten non-durables) plus an estimate of β for durables $\beta^D = 0.17$ and non-durables $\beta^{ND} = 0.76$ (β is assumed constant across industries within each sector). Our theory predicts a positive relationship between labor returns to scale (ν) and short run responses (SRR) of labor to technology shocks that is supported by their evidence. In Table 2 the SRR of labor for 2-digit industries, and standard deviations, are taken directly from Chang and Hong replication files, while the labor returns to scale are proxied by the returns to scale reported in their table 5. Our main result is the levels prediction that short-run responses are positive with increasing returns to labor $\nu > 1$ and negative with decreasing returns to labor $\nu < 1$. The results show that 14 of 20 industries respond the way we would expect,[43] and of the 5 significant (asterisk) responses reported by Chang and Hong all but textile conform to our theory.[44]

Chang and Hong find that there are increasing returns in the majority of industries (14 out of 20) in terms of γ. Estimates of β are both positive: if we combine β with γ, all of the industries have increasing returns so that all of the sectors with a negative or zero short-run impact are inconsistent with our theory: this is 7 industries, meaning 13 are theory consistent. Hence, Chang and Hong's results are broadly supportive of our theoretical result: 13 or 14 of the industries are consistent with our results whether we use γ or

[42] See Basu, Fernald, and Kimball 2006 equation 18, p1424.

[43] This includes *Instruments* which has no short-run response and is the closest estimate to constant returns.

[44] In Appendix A.10 Figure 2 we report the results as a scatter plot.

SIC	Industry	RTS	SRR	SD
23	Apparel	1.24	0.012	0.009
28	Chemicals	1.52	-0.004	0.004
36	Electronic	1.53	-0.009	0.012
34	Fab. Metal	1.29	0.024	0.090
20	Food	0.38	0.001	0.003
25	Furniture	1.18	0.021	0.009*
38	Instruments	0.97	0.000	0.011
31	Leather	0.39	-0.002	0.012
24	Lumber	0.92	-0.028	0.011*
33	Metal	1.29	0.012	0.017
39	Misc	1.41	0.016	0.010*
35	Non-electronic	1.67	0.000	0.013
26	Paper	1.48	0.001	0.008
29	Petrol	0.53	-0.004	0.007
27	Printing	1.49	-0.001	0.008
30	Rubber	1.15	0.022	0.010*
32	Stone	1.36	0.009	0.008
22	Textile	0.86	0.017	0.006*
21	Tobacco	1.08	0.005	0.006
37	Transport	1.12	0.018	0.013

Table 2: Chang and Hong 2006 Results Comparison

$\gamma + \beta$ as our measure of ν.

5 Entry Regulation Shock

We interpret γ in the cost of entry equation (37) as red tape. When red tape increases firm entry costs become more sensitive to the flow of entry. For example, if a resource needed to setup a firm is in inelastic supply, like a government office that provides certificates to enter an industry, then a rise in red tape amplifies congestion. This makes entry more costly, and a firm may wait until a less congested period to attain certification. A ‘deregulatory’ policy decreases γ.[45] Data reported in Figure 1 indicate that red tape, proxied

[45] We adopt the term deregulatory shock following Bilbiie, Ghironi, and Melitz 2007 and authors who interpret entry costs as influenced by regulation (Blanchard and Giavazzi 2003; Poschke 2010; Barseghyan and DiCecio 2011). Whereas these focus on differences in

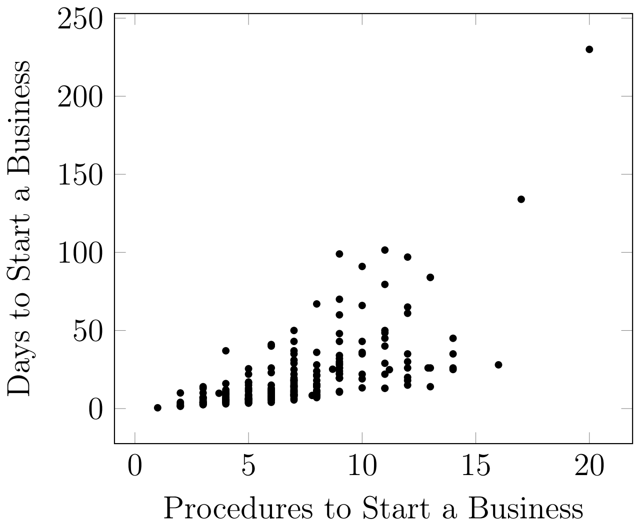

Figure 1: Red Tape and Business Churn

by procedures to start a business, is positively related to the length of time it takes to start a firm which proxies pace of business formation.[46]

Proposition 6. *The economy's speed of adjustment is monotonically decreasing in regulation of business creation.*

The magnitude of the stable root captures the economy's speed of adjustment, as it dictates the speed of adjustment of the sole state variable (number of firms) through the exponential term of (56). Taking the derivative of the stable root, which is negative, with respect to the regulatory parameter gives[47]

$$\Gamma_\gamma = \Gamma_\Delta \Delta_\gamma = \frac{\Delta_\gamma}{(r^2 - 4\Delta)^{\frac{1}{2}}} = \frac{-\Delta}{\gamma(r^2 - 4\Delta)^{\frac{1}{2}}} > 0 \tag{75}$$

The stable root is increasing in the discriminant and the discriminant $\Delta_\gamma =$

fixed exogenous sunk costs and changes in the steady-state stock of operating firms, our interest is endogenous sunk costs and changes in speed of adjustment of firms.

[46]Figure 1 represents 2016 World Bank Doing Business data for 211 countries. Venezuela is the 20 procedures 230 days outlier. New Zealand is the 0.5 days 1 procedure point. Ebell and Haefke 2009 report similar trends in number of procedures and days to start-up for OECD data.

[47]This result is for a given steady-state $\tilde{N}(\bar{\lambda}))$ as γ will also affect $\tilde{N}$ through $\bar{\lambda}$.

$-\frac{\Delta}{\gamma}$ is increasing in the regulatory parameter. Therefore an increase in regulation, increases the the value of the negative root moving it closer to zero and implying slower adjustment. The result implies that economies with less red tape recover faster following a shock.[48] In the context of labor responses to technology shocks, it implies that labor achieves its new steady state faster. The implication that less red tape, helps business churn and aids the dissipation of shocks supports policy work by the IMF and academic literature focused on structural reform in Europe (e.g.di Mauro and Lopez-Garcia 2015).[49]

6 Conclusion

This paper studies the effect of dynamic entry on short-run labor responses to technology shocks. The main insight is that if firm entry is slow to react, then the response of labor to technology shocks will depend on whether labor is employed with decreasing, increasing or constant returns to scale at the firm level. Furthermore the persistence of these deviations will depend on the level of regulation and consequently on the pace of firms' adjustment.

[48]This line of analysis relates to Chatterjee 2005 who focuses on speed of convergence related to capital utilization.

[49]For example, see The Case for Fiscal Policy to Support Structural Reforms (IMF blog, 2017) and Eurozone rebalancing: Are we on the right track for growth? Insights from the CompNet micro-based data (voxEU, Bartelsman, di Mauro, Dorrucci, 2015) on the policy side and Cacciatore, Duval, et al. 2016a; Cacciatore and Fiori 2016; Cacciatore, Duval, et al. 2016b on the academic side.

References

Barseghyan, Levon and Riccardo DiCecio (2011). "Entry costs, industry structure, and cross-country income and TFP differences". In: *Journal of Economic Theory* 146.5, pp. 1828–1851.

— (2016). "Externalities, endogenous productivity, and poverty traps". In: *European Economic Review* 85.C, pp. 112–126.

Basu, Susanto and John Fernald (2001). "Why Is Productivity Procyclical? Why Do We Care?" In: *New Developments in Productivity Analysis*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 225–302.

Basu, Susanto, John G. Fernald, and Miles S. Kimball (2006). "Are technology improvements contractionary?" In: *The American Economic Review* 96.5, pp. 1418–1448.

Berentsen, Aleksander and Christopher Waller (2015). "Optimal Stabilization Policy with Search Externalities". In: *Macroeconomic Dynamics* 19 (03), pp. 669–700.

Bergin, Paul R, Ling Feng, and Ching-Yi Lin (2016). "Firm entry and financial shocks". In: *The Economic Journal*.

Bergin, Paul R. and Ching-Yi Lin (2012). "The dynamic effects of a currency union on trade". In: *Journal of International Economics* 87.2, pp. 191–204.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz (2012). "Endogenous Entry, Product Variety, and Business Cycles". In: *Journal of Political Economy* 120.2, pp. 304–345.

Bilbiie, Florin Ovidiu, Fabio Pietro Ghironi, and Marc J Melitz (2016). "Monopoly Power and Endogenous Product Variety: Distortions and Remedies". In:

Bilbiie, Florin, Fabio Ghironi, and Marc J. Melitz (2007). *Endogenous Entry, Product Variety, and Business Cycles*. NBER Working Papers 13646. National Bureau of Economic Research, Inc.

Blanchard, Olivier and Francesco Giavazzi (2003). “Macroeconomic effects of regulation and deregulation in goods and labor markets”. In: *The Quarterly Journal of Economics* 118.3, pp. 879–907.

Burnside, Craig and Martin Eichenbaum (1996). “Factor-Hoarding and the Propagation of Business-Cycle Shocks”. In: *American Economic Review* 86.5, pp. 1154–1174.

Cacciatore, Matteo, Romain Duval, Giuseppe Fiori, and Fabio Ghironi (2016a). “Market reforms in the time of imbalance”. In: *Journal of Economic Dynamics and Control* 72.C, pp. 69–93.

— (2016b). “Short-term pain for long-term gain: Market deregulation and monetary policy in small open economies”. In: *Journal of International Money and Finance* 68.C, pp. 358–385.

Cacciatore, Matteo and Giuseppe Fiori (2016). “The Macroeconomic Effects of Goods and Labor Marlet Deregulation”. In: *Review of Economic Dynamics* 20, pp. 1–24.

Campbell, Jeffrey R (1998). “Entry, exit, embodied technology, and business cycles”. In: *Review of economic dynamics* 1.2, pp. 371–408.

Cantore, Cristiano, Filippo Ferroni, and Miguel A. Leon-Ledesma (2017). “The dynamics of hours worked and technology”. In: *Journal of Economic Dynamics and Control* 82.C, pp. 67–82.

Caputo, M.R. (2005). *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press.

Chang, Yongsung and Jay H Hong (2006). “Do technological improvements in the manufacturing sector raise or lower employment?” In: *American Economic Review* 96.1, pp. 352–368.

Chatterjee, Santanu (2005). “Capital utilization, economic growth and convergence”. In: *Journal of Economic Dynamics and Control* 29.12, pp. 2093–2124.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson (2003). *What Happens After a Technology Shock?* NBER Working Papers 9819. National Bureau of Economic Research, Inc.

— (2004). “The Response of Hours to a Technology Shock: Evidence Based on Direct Measures of Technology”. In: *Journal of the European Economic Association* 2.2-3, pp. 381–395.

Das, Sanghamitra and Satya P Das (1997). “Dynamics of entry and exit of firms in the presence of entry adjustment costs”. In: *International Journal of Industrial Organization* 15.2, pp. 217–241.

Datta, Bipasa and Huw Dixon (2002). “Technological Change, Entry, and Stock-Market Dynamics: An Analysis of Transition in a Monopolistic Industry”. In: *American Economic Review* 92.2, pp. 231–235.

De Loecker, Jan and Jan Eeckhout (2017). *The rise of market power and the macroeconomic implications*. Tech. rep. National Bureau of Economic Research.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda (2018). *Changing Business Dynamism and Productivity: Shocks vs. Responsiveness*. Working Paper 24236. National Bureau of Economic Research.

Devereux, Michael B, Allen C Head, and Beverly J Lapham (1996). “Aggregate fluctuations with increasing returns to specialization and scale”. In: *Journal of economic dynamics and control* 20.4, pp. 627–656.

di Mauro, Filippo and Paloma Lopez-Garcia (2015). *Assessing European competitiveness: the new CompNet microbased database*. Working Paper Series 1764. European Central Bank.

Ebell, Monique and Christian Haefke (2009). “Product Market Deregulation and the U.S. Employment Miracle”. In: *Review of Economic Dynamics* 12.3, pp. 479–504.

Etro, Federico (2009). *Endogenous market structures and the macroeconomy*. Springer Science & Business Media.

Etro, Federico and Andrea Colciago (2010). “Endogenous Market Structures and the Business Cycle*”. In: *The Economic Journal* 120.549, pp. 1201–1233.

Gali, Jordi (1999). “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” In: *American Economic Review* 89.1, pp. 249–271.

Hamano, Masashige and Francesco Zanetti (2017). “Endogenous product turnover and macroeconomic dynamics”. In: *Review of Economic Dynamics* 26, pp. 263–279.

Hopenhayn, Hugo A. (1992). “Entry, Exit, and Firm Dynamics in Long Run Equilibrium”. In: *Econometrica* 60.5, pp. 1127–1150.

Hornstein, Andreas (1993). “Monopolistic competition, increasing returns to scale, and the importance of productivity shocks”. In: *Journal of Monetary Economics* 31.3, pp. 299–316.

Jaimovich, Nir (2007). “Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations”. In: *Journal of Economic Theory* 137.1, pp. 300–325.

Jaimovich, Nir and Max Floetotto (2008). “Firm dynamics, markup variations, and the business cycle”. In: *Journal of Monetary Economics* 55.7, pp. 1238–1252.

Lewis, Vivien (2009). “Business Cycle Evidence On Firm Entry”. In: *Macroeconomic Dynamics* 13 (05), pp. 605–624.

Lewis, Vivien and Céline Poilly (2012). “Firm entry, markups and the monetary transmission mechanism”. In: *Journal of Monetary Economics* 59.7, pp. 670–685.

Lewis, Vivien and Roland Winkler (2017). “Government spending, entry, and the consumption crowding-in puzzle”. In: *International Economic Review* 58.3, pp. 943–972.

Loualiche, Erik et al. (2014). “Asset Pricing with Entry and Imperfect Competition”. In: *2014 Meeting Papers*. 263. Society for Economic Dynamics.

Mankiw, N Gregory and Michael D Whinston (1986). “Free entry and social inefficiency”. In: *The RAND Journal of Economics*, pp. 48–58.

Melitz, Marc J (2003). “The impact of trade on intra-industry reallocations and aggregate industry productivity”. In: *Econometrica* 71.6, pp. 1695–1725.

Mendoza, Enrique G (1991). “Real business cycles in a small open economy”. In: *The American Economic Review*, pp. 797–818.

Oxborrow, David and Stephen J Turnovsky (2017). “Closing the small open economy model: A demographic approach”. In: *Review of International Economics* 25.1, pp. 44–75.

Poschke, Markus (2010). “The Regulation of Entry and Aggregate Productivity”. In: *Economic Journal* 120.549, pp. 1175–1200.

Poutineau, Jean-Christophe and Gauthier Vermandel (2015). “Financial frictions and the extensive margin of activity”. In: *Research in Economics* 69.4, pp. 525–554.

Rebei, Nooman (2014). “What (really) accounts for the fall in hours after a technology shock?” In: *Journal of Economic Dynamics and Control* 45, pp. 330–352.

Schmitt-Grohé, Stephanie and Martın Uribe (2003). “Closing small open economy models”. In: *Journal of International Economics* 61.1, pp. 163–185.

Sen, Partha and Stephen J. Turnovsky (1990). “Investment tax credit in an open economy”. In: *Journal of Public Economics* 42.3, pp. 277–299.

Seoane, Hernán D (2015). “Near unit root small open economies”. In: *Journal of Economic Dynamics and Control* 53, pp. 37–46.

Stokey, Nancy L (2008). *The Economics of Inaction: Stochastic Control models with fixed costs*. Princeton University Press.

Turnovsky, S.J. (1997). *International Macroeconomic Dynamics*. MIT Press.

Turnovsky, Stephen J (2002). “Knife-edge conditions and the macrodynamics of small open economies”. In: *Macroeconomic Dynamics* 6.2, pp. 307–335.

Uribe, M. and S. Schmitt-Grohé (2017). *Open Economy Macroeconomics*. Princeton University Press.

Wang, Pengfei and Yi Wen (2011). “Understanding the Effects of Technology Shocks”. In: *Review of Economic Dynamics* 14.4, pp. 705–724.

Young, Andrew T (2004). “Labor’s share fluctuations, biased technical change, and the business cycle”. In: *Review of Economic Dynamics* 7.4, pp. 916–931.

A Appendix

A.1 Household Optimization

The Hamiltionian and optimality conditions are

$$\hat{\mathcal{H}}(t) = U(C,H) + \lambda(t)[rB + wH + \Pi - C - G] \tag{76}$$

$$\begin{aligned}
\hat{\mathcal{H}}_C = 0: &\quad U_C(C) - \lambda = 0 && (77)\\
\hat{\mathcal{H}}_H = 0: &\quad U_H(H) + \lambda w = 0 && (78)\\
\hat{\mathcal{H}}_B = \rho\lambda - \dot{\lambda}: &\quad \lambda r = \rho\lambda - \dot{\lambda} && (79)\\
\hat{\mathcal{H}}_\lambda = \dot{B}: &\quad \dot{B} = rB + wH + \Pi - C - G && (80)
\end{aligned}$$

The presence of a small open economy and international capital markets $\rho = r$ means that the household can completely smooth its consumption so (79) implies $\dot{\lambda} = 0$. Therefore marginal utility of wealth is unchanging over time. $\lambda = \bar{\lambda}$ combined with additively separable preferences $u_{CH} = 0$ this implies from (77) that consumption is constant and in a one-one relationship with marginal utility of wealth.[50]

$$\bar{C} = C(\bar{\lambda}) \tag{81}$$

This relationship from (77) then implies labor only varies with real wage from (78)

$$H = H(\bar{\lambda}, w) = H(\bar{C}, w) \tag{82}$$

This represents the households labor supply.

[50]We could not make the final step from (77) is $u_{CH} \neq 0$. Imposing additive separability and therefore constant consumption, we simplify analysis of dynamics as C can be treated as fixed.

A.2 General Equilibrium Effect of Entry on Output

There are two ways to think of the effect of an entrant on aggregate output $\frac{dY}{dN}$, and they offer different intuitions. The first begins with $Y = Ny$ and the second begin with $Y = AN^{1-\nu}H^{\nu} - N\phi$.

1. $\frac{dY(N,y(N,H))}{dN} = \frac{d[Ny]}{dN} = y + N\frac{dy}{dN}$ An entrant always causes 'business stealing' from other firms: a fall in output at the firm level or analogously, by (24), a fall in an each incumbents' profits.

$$\frac{dy}{dN} < 0 \tag{83}$$

$$\frac{dy}{dN} = \frac{d\left(AN^{-\nu}H^{\nu} - \phi\right)}{dN} \tag{84}$$

$$= -\nu\frac{(y+\phi)}{N} + \nu\frac{(y+\phi)}{H}\frac{dH}{dN} \tag{85}$$

$$= \nu\frac{(y+\phi)}{N}\left[\varepsilon - 1\right] < 0 \tag{86}$$

$$= Y_H\frac{h}{N}\left[\varepsilon - 1\right] \tag{87}$$

Therefore the aggregate business stealing effect is

$$N\frac{dy}{dN} = \nu(y+\phi)(\varepsilon - 1) \tag{88}$$

This also implies the effect on operating profits is negative and less than proportional

$$\frac{d\pi}{dN} = \left(1 - \frac{\nu}{\mu}\right)\frac{dy}{dN} < 0 \tag{89}$$

At the aggregate level it is not clear whether the negative business stealing effect of an entrant aggregated across all incumbents offsets the positive effect of the new firms' extra output.

$$\frac{dY}{dN} = \frac{d(Ny)}{dN} \tag{90}$$
$$= y + N\frac{dy}{dN} \tag{91}$$
$$= y + \nu Ah^{\nu}(\varepsilon - 1) \tag{92}$$
$$= Ah^{\nu}(1 - (1 - \varepsilon)\nu) - \phi \tag{93}$$
$$= \frac{(1-\nu)(1+\eta)}{1+\eta-\nu}Ah^{\nu} - \phi \tag{94}$$
$$= \varepsilon(1+\eta)Ah^{\nu} - \phi \tag{95}$$

The final representation makes clear the crucial effect of returns to scale. It reads that an entrant has a negative effect by bringing in an extra fixed cost, but it has another positive negative or zero effect depending on ε.

2. Alternatively use (34), where the first term is the partial derivative effect of an entrant which we have explained is ambiguous based on ν, and the second term is the labor response which is also ambiguous based on ν.

$$\frac{dY(N,H)}{dN} = \frac{d[AN^{1-\nu}H^{\nu} - N\phi]}{dN} = Y_N + Y_H H_N \tag{96}$$
$$= \pi - \left(1 - \frac{1}{\mu}\right) Y_H \frac{H}{N} + Y_H H_N \tag{97}$$
$$= \pi - \left(1 - \frac{1}{\mu} - \varepsilon\right) Y_H h \tag{98}$$

A.3 Bonds

The dynamic equation (41c) is a first-order, linear, nonhomogeneous ordinary differential equation in B. Rewrite in standard form

$$\dot{B} - rB = Y - \frac{q^2}{2\gamma} - C - G \tag{99}$$

Multiply by the integrating factor e^{-rt}

$$e^{-rt}\dot{B} - re^{-rt}B = e^{-rt}\left[Y - \frac{q^2}{2\gamma} - C - G\right] \tag{100}$$

Notice the left-hand side as the result of a product rule differentiation, and use this to help integrate

$$e^{-rt}B = \kappa + \int_0^\infty e^{-rt}\left[Y - \frac{q^2}{2\gamma} - C - G\right] dt \tag{101}$$

To find the constant of integration κ, evaluate at $t = 0$ and use the initial condition $B(0) = B_0$

$$B(0) = \kappa = B_0 \tag{102}$$

Substitute this back in (101), then evaluate at $t \to \infty$. Use the transversality condition (9) which makes the left-hand side zero as $\lambda = \bar{\lambda}$. Therefore

$$0 = B_0 + \int_0^\infty e^{-rt}\left[Y - \frac{q^2}{2\gamma} - C - G\right] dt \tag{58}$$

A.4 Profit Maximization with Variable Returns to Scale

$$\max_h \pi_i = p_i y_i - wh_i \tag{103}$$

$$\text{s.t.} \quad \frac{p_i}{P} = \left(\frac{Y}{N^\varsigma y_i}\right)^{\frac{1}{\theta}} \tag{11}$$

$$y_i = Ah_i^\nu - \phi \tag{12}$$

$$\pi = \left(\frac{Y}{N^{\varsigma}}\right)^{\frac{1}{\theta}} (Ah^{\nu} - \phi)^{1-\frac{1}{\theta}} - wh \tag{104}$$

$$\pi_h = \left(\frac{Y}{N^{\varsigma}}\right)^{\frac{1}{\theta}} \left(1 - \frac{1}{\theta}\right) (Ah^{\nu} - \phi)^{-\frac{1}{\theta}} \cdot A\nu h^{\nu-1} - w \tag{105}$$

$$\pi_{hh} = -\frac{1}{\theta}\frac{(\pi_h + w)}{y}\frac{(y+\phi)\nu}{h} + \frac{(\pi_h + w)(\nu - 1)}{h} \tag{106}$$

$$= \frac{\pi_h + w}{h}\left[\nu\left(1 - \frac{1}{\theta} - \frac{\phi}{\theta y}\right) - 1\right] \tag{107}$$

$$\pi_{hh} < 0 \quad \iff \quad 1 - \frac{1 + s_{\phi}}{\theta} < \frac{1}{\nu} \tag{108}$$

The second-order condition $\pi_{hh} < 0$ for maximization is always satisfied when $\nu \leq 1$. However with increasing returns $\nu > 1$ it is possible that the term in square brackets is positive, hence there is a necessary and sufficient condition, which can be expressed as

$$1 - \frac{1 + s_{\phi}}{\theta} < \frac{1}{\nu} \tag{109}$$

$$\theta\left(\frac{\theta - 1}{\theta} - \frac{1}{\nu}\right) < s_{\phi} \tag{110}$$

$$\left(\frac{1}{\mu} - \frac{1}{\nu}\right) < \frac{s_{\phi}}{\theta} \tag{111}$$

Throughout the paper we impose that the markup $\frac{\theta}{\theta-1} \equiv \mu$ exceeds returns to scale $\mu > \nu$ (this is necessary for a well-defined steady-state), but it is also a sufficient condition for the second-order condition to hold since $\frac{1}{\mu} - \frac{1}{\nu} < 0 < \frac{\phi}{\theta y}$. Under perfect competition $\theta \to \infty$ so $\mu \to 1$, there must be increasing marginal costs $\nu < 1$ which gives the outcome that with a fixed cost, Walrasian equilibrium only exists with increasing marginal costs, where marginal cost intersect minimum average costs at a firm's *minimum efficient scale.*

To find the profit maximizing ($\pi_h = 0$) outcome, exploit symmetry $Y/N =$

y.

$$\pi_h = \left(\frac{Y}{N^{\varsigma} y}\right)^{\frac{1}{\theta}} \left(1 - \frac{1}{\theta}\right) \cdot A\nu h^{\nu-1} - w = 0 \tag{112}$$

$$\pi_h = N^{\frac{1-\varsigma}{\theta}} \left(1 - \frac{1}{\theta}\right) \cdot A\nu h^{\nu-1} - w = 0 \tag{113}$$

$$\pi_h = N^{\frac{1-\varsigma}{\theta}} \left(\frac{\theta - 1}{\theta}\right) \nu A h^{\nu-1} - w = 0 \tag{114}$$

where we ignore love of variety by assuming $\varsigma = 1$.

A.5 General Equilibrium Labor Behavior

Proof of Proposition 2. Take the derivative of (27)

$$H_N = \frac{1-\nu}{1+\eta-\nu} \frac{H}{N} \tag{115}$$

Therefore, the elasticity follows naturally

$$\varepsilon = H_N \frac{N}{H} = \frac{1-\nu}{1+\eta-\nu} \tag{116}$$

The elasticity is less than 1, it approaches 1 in the indivisible labor limit.

$$\lim_{\eta \to 0} \varepsilon = 1 \tag{117}$$

$$\lim_{\eta \to \infty} \varepsilon = \begin{cases} 0^{+} & \nu < 1 \\ 0^{-} & \nu > 1 \end{cases} \tag{118}$$

□

A.6 Detailed Jacobian

The Jacobian matrix of the 3-dimensional system is as follows (all elements are evaluated at steady state)

$$\mathbf{J} = \left[\left[\begin{array}{ccc} 0 & \frac{d\dot{N}}{dq} & 0 \\ \frac{d\dot{q}}{dN} & \frac{d\dot{q}}{dq} & 0 \\ \frac{d\dot{B}}{dN} & \frac{d\dot{B}}{dq} & \frac{d\dot{B}}{dB} \end{array}\right]\right|^{\tilde{}} = \left[\begin{array}{ccc} 0 & \frac{1}{\gamma} & 0 \\ -\frac{d\tilde{\pi}}{dN} & r & 0 \\ \frac{d\tilde{Y}}{dN} & -\frac{d\tilde{\mathcal{C}}}{dq} & r \end{array}\right] \tag{119}$$

where,

$$\frac{d\tilde{\mathcal{C}}}{dq} = \frac{\tilde{q}}{\gamma} \tag{120}$$

$$\frac{d\tilde{\pi}}{dN} = \frac{\tilde{\pi} + \phi}{\tilde{N}(\bar{\lambda})}\left(\frac{-\eta\nu}{1+\eta-\nu}\right) \tag{121}$$

$$\frac{d\tilde{Y}}{dN} = A\tilde{h}^{\nu}\left(1+\nu\left(\frac{1-\tilde{h}}{\tilde{h}}\right)\right) - \phi \tag{122}$$

where $\tilde{q} = \tilde{\pi} = 0$ (from (42) and (43)) and (44) gives $\tilde{h}$ as a function of exogenous parameters, but $\tilde{N}(\bar{\lambda})$ depends on endogenously determined steady-state consumption index given in (47). Section 2.3.1 and equation (35) help with these derivations, and make clear that both π an Y responses depend on business stealing $\frac{d\tilde{y}}{dN}$.

A.7 Jacobian Results

In the results that follow, the trace, determinant, eigenvalue relationships are useful

$$\Delta = \Gamma\Gamma^{U} \tag{123}$$

$$r = \Gamma + \Gamma^{U} \tag{124}$$

$$\Delta = \Gamma(r-\Gamma) \tag{125}$$

$$(r^2 - 4\Delta)^{\frac{1}{2}} = r - 2\Gamma \tag{126}$$

The determinant of the entry subsystem $\det(\mathbf{B}) = \Delta(\tilde{N}(\bar{\lambda}))$ is increasing in $\bar{\lambda}$.

$$\Delta_\lambda = \Delta_N \tilde{N}_\lambda = -\frac{\Delta}{\tilde{N}} \cdot \frac{\tilde{N}}{\eta\bar{\lambda}} = -\frac{\Delta}{\eta\bar{\lambda}} > 0 \tag{127}$$

The stable root is increasing in the determinant

$$\Gamma_\Delta = -\frac{r}{2}\left(\frac{1}{2}\left(1 - \frac{4\Delta}{r^2}\right)^{\frac{-1}{2}} \cdot \frac{-4}{r^2}\right) \tag{128}$$

$$= \frac{1}{(r^2 - 4\Delta)^{\frac{1}{2}}} = \frac{1}{r - 2\Gamma} > 0 \tag{129}$$

and therefore increasing in the number of firms

$$\frac{d\Gamma}{d\tilde{N}} = \Gamma_\Delta \Delta_N = \frac{\Gamma(\Gamma - r)}{r - 2\Gamma}\frac{1}{\tilde{N}} > 0 \tag{130}$$

Therefore the stable root is increasing in $\bar{\lambda}$

$$\Gamma_{\bar{\lambda}} = \Gamma_\Delta \Delta_\lambda = \Gamma_\Delta \Delta_{\tilde{N}} \tilde{N}_\lambda > 0 \tag{131}$$

which proves Lemma 3.

This can be written

$$\Gamma_{\bar{\lambda}} = -\frac{\Delta}{\eta\bar{\lambda}(r^2 - 4\Delta)^{\frac{1}{2}}} = \frac{1}{\eta\bar{\lambda}}\frac{\Gamma(\Gamma - r)}{r - 2\Gamma} > 0$$

A.8 Steady-state Proofs

Proof of Proposition 3.

$$\tilde{\Omega} = \left(\varepsilon - 1 + \frac{1}{\mu}\right) Y_H \tilde{h}$$

$$\tilde{\Omega} = \nu \frac{\phi}{1 - \frac{\nu}{\mu}}\left(\varepsilon - 1 + \frac{1}{\mu}\right) = \frac{\nu\phi\mu}{\mu - \nu}\left(\frac{1}{\mu} - \frac{\eta}{1 + \eta - \nu}\right)$$

$$\operatorname{sgn}\tilde{\Omega} = \operatorname{sgn}\left[\varepsilon - \left(\frac{\mu - 1}{\mu}\right)\right]$$

where $\operatorname{sgn}\varepsilon = \operatorname{sgn}(1 - \nu)$ since $\varepsilon = \frac{1-\nu}{1+\eta-v}$ from (32). □

Repeating the steady-state bond condition here

$$\tilde{B}(\bar{\lambda}, A) = B_0 - \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}(N_0 - \tilde{N}(\bar{\lambda})) \tag{62}$$

The total derivative of steady-state bonds with respect to inverse consumption is

$$\frac{d\tilde{B}}{d\bar{\lambda}} = -\tilde{\Omega}\left(\frac{d\left(\frac{N_0 - \tilde{N}(\bar{\lambda})}{\Gamma(\tilde{N}(\bar{\lambda}))-r}\right)}{d\bar{\lambda}}\right) = \tilde{\Omega}\left[\frac{(\Gamma(\bar{\lambda}) - r)\frac{d\tilde{N}}{d\bar{\lambda}} + [N_0 - \tilde{N}(\bar{\lambda})]\frac{d\Gamma(\tilde{N})}{d\bar{\lambda}}}{(\Gamma(\bar{\lambda}) - r)^2}\right] \tag{132}$$

The response of steady-state bonds to inverse consumption $\bar{\lambda}$ is ambiguous because both $\tilde{\Omega}$ and $[N_0 - \tilde{N}(\bar{\lambda})]$ are ambiguously signed. Since this model is path-dependent (steady-state depends on initial conditions $\tilde{N}(\bar{\lambda}, N_0)$ due to (62)), we cannot evaluate at $N_0 = \tilde{N}$, which removes the changing eigenvalue effect (see Caputo 2005, p. 475-477 for this common approach).[51] Instead we follow Turnovsky 1997, p.68 (footnote 8) and assume this component $[N_0 - \tilde{N}]$ is small, which – to a linear approximation – removes the changing eigenvalue effect.

Lemma 5. *The effect of a change in the consumption index on bonds is*

$$\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\left[\frac{\Gamma}{r - 2\Gamma}\left(\frac{r}{\Gamma} - 3 + \frac{N_0}{\tilde{N}}\right)\right]\frac{\tilde{N}}{\bar{\lambda}\eta} \tag{133}$$

Proof. From (62) a change in consumption index only affects steady-state bonds indirectly through its effect on steady-state stock of firms

$$\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{d\tilde{B}}{d\tilde{N}}\frac{d\tilde{N}}{d\bar{\lambda}} \tag{134}$$

Then steady-state stock of firms affects bonds directly $\frac{\partial \tilde{B}}{\partial \tilde{N}}$ through $\tilde{N}$ and

[51] Attempting this approach here introduces another fixed point problem since changing N_0 to equal $\tilde{N}$ will in turn change $\tilde{N}$ due to path-dependency.

indirectly $\frac{d\tilde{B}}{d\Gamma}\frac{d\Gamma}{d\tilde{N}}$ through the eigenvalue $\Gamma(\tilde{N}(\bar{\lambda}))$:

$$\frac{d\tilde{B}}{d\tilde{N}} = \frac{\partial \tilde{B}}{\partial \tilde{N}} + \frac{d\tilde{B}}{d\Gamma}\frac{d\Gamma}{d\tilde{N}} = \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\left[1 + \left(\frac{N_0 - \tilde{N}(\bar{\lambda})}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\right)\frac{d\Gamma}{d\tilde{N}}\right] \tag{135}$$

Therefore the effect of a change in consumption index on bonds through eigenvalues is an indirect-indirect effect.

$$\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{d\tilde{B}}{d\tilde{N}}\frac{d\tilde{N}}{d\bar{\lambda}} = \left(\frac{\partial \tilde{B}}{\partial \tilde{N}} + \frac{d\tilde{B}}{d\Gamma}\frac{d\Gamma}{d\tilde{N}}\right)\frac{d\tilde{N}}{d\bar{\lambda}} \tag{136}$$

$$= \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\left[1 + \left(\frac{N_0 - \tilde{N}(\bar{\lambda})}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\right)\frac{d\Gamma}{d\tilde{N}}\right]\frac{d\tilde{N}}{d\bar{\lambda}} \tag{137}$$

Using (130) the term in square brackets simplifies

$$\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r}\left[\frac{\Gamma}{r - 2\Gamma}\left(\frac{r}{\Gamma} - 3 + \frac{N_0}{\tilde{N}}\right)\right]\frac{d\tilde{N}}{d\bar{\lambda}} \tag{138}$$

Therefore substituting in (159) gives (133). □

Corollary 2. *If* $\frac{N_0}{\tilde{N}(\bar{\lambda})} < 3 - \frac{r}{\Gamma}$ *then*

$$\operatorname{sgn}\frac{d\tilde{B}}{d\bar{\lambda}} = -\operatorname{sgn}\tilde{\Omega} \tag{139}$$

Proof. From (133) this result ensures the term in curled parenthesis is negative. □

Hence a sufficient condition is $\frac{N_0}{\tilde{N}} < 3$, which allows for both entry and exit $-\tilde{N} < N_0 - \tilde{N} < 2\tilde{N}$. The economic interpretation is that the initial stock of firms (market size) is greater than zero and less than three times the steady-state stock of firms. This is more general than the (commonly assumed) stronger condition that the initial condition is arbitrarily close to steady state $\frac{N_0}{\tilde{N}} \to 1$. This condition simply ensures we ignore the changing eigenvalue effect.

Corollary 3. *If* $[N_0 - \tilde{N}(\bar{\lambda})] \to 0$ *then*

$$\operatorname{sgn} \frac{d\tilde{B}}{d\bar{\lambda}} = -\operatorname{sgn} \tilde{\Omega} \tag{140}$$

Proof. From (135) as $N_0 - \tilde{N}(\bar{\lambda}) \to 0$

$$\frac{d\tilde{B}}{d\tilde{N}} \approx \frac{\partial \tilde{B}}{\partial \tilde{N}} = \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r} \tag{141}$$

$$\frac{d\tilde{B}}{d\bar{\lambda}} \approx \frac{\partial \tilde{B}}{\partial \tilde{N}} \frac{d\tilde{N}}{d\bar{\lambda}} = \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(\bar{\lambda})) - r} \frac{\tilde{N}}{\bar{\lambda}\eta} \tag{142}$$

□

Lemma 6 (Steady-state Existence)**.** *By the intermediate-value theorem at least one steady-state solution exists.*

Proof of Lemma 6. Split the steady-state excess demand function into two functions: an income function $f(\bar{\lambda}) = \tilde{w}\tilde{H}(\bar{\lambda}) + rB(\bar{\lambda})$ and an expenditure function $g(\bar{\lambda}) = C(\bar{\lambda}) + G$, so we have $f(\bar{\lambda}) - g(\bar{\lambda}) = 0$. Analyze the functions for the limits of $\bar{\lambda}$. Existence follows from the functional forms for $H(\bar{\lambda}, A) = (\bar{\lambda}w)^{\frac{1}{\eta}}$ and $C(\bar{\lambda}) = \frac{1}{\bar{\lambda}}$. Also that $\tilde{B}$ is bounded in (62) since $\tilde{N}$ is bounded as it is proportional to $\tilde{H}$, which lies in $[0, 1]$. $\lim_{\lambda \to 0} H = 0$ and $\lim_{\lambda \to 0} C = \infty$ so expenditure exceeds income. $\lim_{\lambda \to \infty} H = 1$ and $\lim_{\lambda \to \infty} C = 0$, so income exceeds expenditure. Hence for at least one intermediate value of λ (63) is satisfied. □

Proof of Lemma 4. We aim to show

$$\tilde{w}\frac{d\tilde{H}}{d\lambda} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\tilde{C}}{d\bar{\lambda}} > 0 \tag{64}$$

Since $\frac{d\tilde{C}}{d\lambda} < 0$, a sufficient condition is to show that $\tilde{w}\frac{d\tilde{H}}{d\lambda} + r\frac{d\tilde{B}}{d\lambda} > 0$. That is, we show that the positive labor effect always dominates the (potentially) negative bond effect.

$$\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{\tilde{Y}_H}{\mu}\frac{d\tilde{H}}{d\bar{\lambda}} + r\tilde{\Omega}\left[\frac{(\Gamma - r)\frac{d\tilde{N}}{d\lambda} + [N_0 - \tilde{N}]\frac{d\Gamma}{d\lambda}}{(\Gamma - r)^2}\right] \tag{143}$$

Substitute $\tilde{\Omega} = \left(\varepsilon - 1 + \frac{1}{\mu}\right) \tilde{Y_H} \tilde{h}$ and $\frac{d\tilde{N}}{d\bar{\lambda}} = \frac{d\tilde{H}}{d\bar{\lambda}} \frac{1}{\tilde{h}}$

$$= \left[\frac{Y_H}{\mu} \frac{d\tilde{H}}{d\bar{\lambda}} (\Gamma - r) + r \left(\varepsilon - 1 + \frac{1}{\mu} \right) Y_H \frac{d\tilde{H}}{d\bar{\lambda}} + \frac{r \left(\varepsilon - 1 + \frac{1}{\mu}\right) Y_H \tilde{h} (N_0 - \tilde{N})}{\Gamma - r} \frac{d\Gamma}{d\bar{\lambda}} \right] \frac{1}{\Gamma - r} \tag{144}$$

$$= \left[\frac{1}{\mu} (\Gamma - r) + r \left(\varepsilon - 1 + \frac{1}{\mu} \right) + \frac{r \left(\varepsilon - 1 + \frac{1}{\mu}\right) \tilde{h} (N_0 - \tilde{N})}{(\Gamma - r) \frac{d\tilde{H}}{d\bar{\lambda}}} \frac{d\Gamma}{d\bar{\lambda}} \right] \frac{Y_H \frac{d\tilde{H}}{d\bar{\lambda}}}{\Gamma - r} \tag{145}$$

Cancel $\frac{r}{\mu}$ and use that $\frac{d\tilde{H}}{d\bar{\lambda}} = \frac{d\tilde{N}}{d\bar{\lambda}} \tilde{h}$

$$= \left[\frac{1}{\mu} \Gamma + r (\varepsilon - 1) + \frac{r \left(\varepsilon - 1 + \frac{1}{\mu}\right) (N_0 - \tilde{N})}{\Gamma - r} \frac{\frac{d\Gamma}{d\bar{\lambda}}}{\frac{d\tilde{N}}{d\bar{\lambda}}} \right] \frac{Y_H \frac{d\tilde{H}}{d\bar{\lambda}}}{\Gamma - r} \tag{146}$$

Remembering $\varepsilon - 1 < 0$, the first two terms are negative and the third term (the changing eigenvalue term $\frac{d\Gamma}{d\bar{\lambda}}$) is ambiguous. As with signing $\tilde{B}_{\bar{\lambda}}$, a sufficient condition to remove the problematic changing eigenvalue term is $N_0 - \tilde{N} \to 0$. Although a weaker, but messier, sufficient condition is:

$$\left(\varepsilon - 1 + \frac{1}{\mu}\right) \left(\frac{N_0}{\tilde{N}} - 1\right) \frac{\Gamma}{r - 2\Gamma} \leq - \left(\frac{\Gamma}{r\mu} + \varepsilon - 1\right) \tag{147}$$

$$\left(\varepsilon - 1 + \frac{1}{\mu}\right) \left(\frac{N_0}{\tilde{N}} - 1\right) \geq - \left(\frac{\varepsilon - 1}{\Gamma} + \frac{1}{r\mu}\right) (r - 2\Gamma) \tag{148}$$

The right-hand side is negative so this condition always holds if there is entry $N_0 < \tilde{N}$ and $\varepsilon - 1 + \frac{1}{\mu} < 0$ implying $\tilde{\Omega} < 0$. Or if there is exit $N_0 > \tilde{N}$ and and $\varepsilon - 1 + \frac{1}{\mu} > 0$ implying $\tilde{\Omega} > 0$.

□

A.9 Dynamics

Rather than defining steady-state as a function of $\tilde{h}(A), \tilde{w}(A)$ as in (47) and (48), since both depend on A and we are investigating changes in A it is useful substitute out. Repeating $\tilde{B}$, expressing dependence on A, is also useful. A only affects $\tilde{B}$ through $\tilde{N}$, which it affects directly and indirectly: $\tilde{N}(A, \bar{\lambda}(A))$ via (149).

$$\tilde{N}(\bar{\lambda}, A) = \left(\bar{\lambda}\frac{\nu}{\mu}\right)^{\frac{1}{\eta}} A^{\frac{1+\eta}{\nu\eta}} \left(\frac{\mu-\nu}{\mu\phi}\right)^{\frac{1+\eta-\nu}{\nu\eta}} \tag{149}$$

$$\tilde{H}(\bar{\lambda}, A) = \tilde{h}(A)\tilde{N}(\bar{\lambda}, A) = \left(\bar{\lambda}\frac{\nu}{\mu}\right)^{\frac{1}{\eta}} A^{\frac{1}{\nu\eta}} \left(\frac{\mu-\nu}{\mu\phi}\right)^{\frac{1-\nu}{\nu\eta}} \tag{150}$$

$$\tilde{B}(\tilde{N}(A, \bar{\lambda}(A))) = B_0 - \frac{\tilde{\Omega}}{\Gamma(\tilde{N}(A, \bar{\lambda}(A))) - r}(N_0 - \tilde{N}(\tilde{N}(A, \bar{\lambda}(A))) \tag{62}$$

Technology change has a direct (partial) and an indirect (consumption) effect on the core endogenous model variables

$$\frac{dX}{dA} = \frac{\partial X}{\partial A} + \frac{dX}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA}, \quad X \in \{\bar{C}, \tilde{N}, \tilde{B}\} \tag{151}$$

The direct (partial) effects of A holding $\bar{\lambda}$ constant are simple to calculate. There is no partial effect on consumption, only an indirect effect.

$$\frac{\partial \bar{C}}{\partial A} = 0 \tag{152}$$

$$\frac{\partial \tilde{N}}{\partial A} = \frac{(1+\eta)\tilde{N}}{\nu\eta A} > 0 \tag{153}$$

$$\frac{\partial \tilde{B}}{\partial A} \approx \frac{\tilde{\Omega}}{\Gamma - r}\frac{\partial \tilde{N}}{\partial A} \gtreqless 0 \implies \operatorname{sgn}\frac{\partial \tilde{B}}{\partial A} = \operatorname{sgn} -\tilde{\Omega} \tag{154}$$

$$\frac{\partial \tilde{H}}{\partial A} = \frac{\tilde{H}}{\nu A\eta} > 0 \tag{155}$$

From the steady state market clearing condition (63), we can use the implicit function theorem to infer that technology decreases the marginal utility of consumption and therfore increase consumption (since through (7) consumption and marginal utility are inversely related).

Proposition 7 (Technology Effect on Steady-state Consumption)**.**

$$\frac{d\bar{\lambda}}{dA} < 0 \tag{156}$$

$$\frac{d\bar{C}}{dA} = \frac{d\bar{C}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} > 0 \tag{157}$$

$$\frac{d\bar{C}}{d\bar{\lambda}} = -\frac{1}{\bar{\lambda}^2} < 0 \tag{158}$$

Therefore an increase in technology increases consumption (decreases marginal utility), which, from (47) and (48), will have an indirect effect of decreasing numbers of firms and labor. This is because consumption crowds out investment in firms.

$$\frac{d\tilde{N}}{d\bar{\lambda}} = \frac{\tilde{N}}{\eta\bar{\lambda}} > 0 \tag{159}$$

$$\frac{d\tilde{B}}{d\bar{\lambda}} = \frac{d\tilde{B}}{d\tilde{N}}\frac{d\tilde{N}}{d\bar{\lambda}} \approx \frac{\tilde{\Omega}}{\Gamma - r}\frac{d\tilde{N}}{d\bar{\lambda}} \implies \operatorname{sgn}\frac{d\tilde{B}}{d\bar{\lambda}} = -\operatorname{sgn}\tilde{\Omega} \tag{160}$$

$$\frac{d\tilde{H}}{d\bar{\lambda}} = \tilde{h}\frac{d\tilde{N}}{d\bar{\lambda}} = \frac{\tilde{H}}{\eta\bar{\lambda}} > 0 \tag{161}$$

Proof of Proposition 7. The total derivative of (63) with respect to technology is

$$\frac{d\tilde{w}}{dA}\tilde{H} + \tilde{w}\left(\frac{\partial\tilde{H}}{\partial A} + \frac{d\tilde{H}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA}\right) + r\left(\frac{\partial\tilde{B}}{\partial A} + \frac{d\tilde{B}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA}\right) - \frac{dC}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} = 0 \tag{162}$$

Therefore

$$\frac{d\bar{\lambda}}{dA} = -\frac{\frac{d\tilde{w}}{dA}\tilde{H} + \tilde{w}\frac{\partial\tilde{H}}{\partial A} + r\frac{\partial\tilde{B}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}} < 0 \tag{163}$$

The denominator is positive under sufficient condition (65) or stronger sufficient condition $N_0 - \tilde{N} \to 0$. Let's focus on the numerator

$$\frac{d\tilde{w}}{dA}\tilde{H} + \tilde{w}\frac{\partial \tilde{H}}{\partial A} + r\frac{\partial \tilde{B}}{\partial A} \tag{164}$$

which appears to be ambiguous. We shall show it is positive implying (163) is negative.

$$\frac{d\tilde{w}}{dA}\tilde{H} + \tilde{w}\frac{\partial \tilde{H}}{\partial A} + r\frac{\partial \tilde{B}}{\partial A} \tag{165}$$

$$= \frac{\tilde{w}}{\nu A}\tilde{H} + \tilde{w}\frac{\tilde{H}}{\nu A\eta} + r\frac{\tilde{\Omega}}{\Gamma - r}\frac{(1+\eta)\tilde{N}}{\nu\eta A} = \frac{1+\eta}{\nu A}\left[\frac{\tilde{w}\tilde{H}}{(1+\eta)} + \frac{\tilde{w}\tilde{H}}{(1+\eta)\eta} + r\frac{\tilde{\Omega}}{\Gamma - r}\frac{\tilde{N}}{\eta}\right] \tag{166}$$

$$= \frac{1+\eta}{\nu A}\left[\frac{\tilde{w}\tilde{H}}{\eta} + r\frac{\tilde{\Omega}}{\Gamma - r}\frac{\tilde{N}}{\eta}\right] = \frac{1+\eta}{\nu A}\left[\frac{\frac{\tilde{Y_H}}{\mu}\tilde{H}}{\eta} + r\frac{\tilde{\Omega}}{\Gamma - r}\frac{\tilde{N}}{\eta}\right] \tag{167}$$

Substitute $\tilde{\Omega} = (\varepsilon - 1 + \frac{1}{\mu})\tilde{Y_H}\frac{\tilde{H}}{\tilde{N}}$

$$= \frac{1+\eta}{\nu A}\left[\frac{\frac{\tilde{Y_H}}{\mu}\tilde{H}}{\eta} + r\frac{(\varepsilon - 1 + \frac{1}{\mu})\tilde{Y_H}\frac{\tilde{H}}{\tilde{N}}}{\Gamma - r}\frac{\tilde{N}}{\eta}\right] = \frac{(1+\eta)\tilde{Y_H}\tilde{H}}{\nu A\eta}\left[\frac{1}{\mu} + r\frac{(\varepsilon - 1 + \frac{1}{\mu})}{\Gamma - r}\right] \tag{168}$$

$$= \frac{(1+\eta)\tilde{Y_H}\tilde{H}}{\nu A\eta}\frac{1}{(\Gamma - r)}\left[\frac{\Gamma}{\mu} + r(\varepsilon - 1)\right] = \frac{(1+\eta)\tilde{N}(\tilde{y} + \phi)}{A\eta}\frac{1}{(\Gamma - r)}\left[\frac{\Gamma}{\mu} + r(\varepsilon - 1)\right] > 0 \tag{169}$$

Using $\frac{\tilde{H}}{\eta\bar{\lambda}} = \frac{d\tilde{H}}{d\bar{\lambda}}$ we can show

$$= \frac{(1+\eta)\bar{\lambda}}{\nu A}\frac{\tilde{Y_H}\frac{d\tilde{H}}{d\bar{\lambda}}}{(\Gamma - r)}\left[\frac{\Gamma}{\mu} + r(\varepsilon - 1)\right] \tag{170}$$

Substitute (146) (ignore changing eigenvalue effect)

$$= \frac{(1+\eta)\bar{\lambda}}{\nu A}\left(\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}\right) > 0 \tag{171}$$

Therefore

$$\frac{d\bar{\lambda}}{dA} = -\frac{\frac{d\tilde{w}}{dA}\tilde{H} + \tilde{w}\frac{\partial \tilde{H}}{\partial A} + r\frac{\partial \tilde{B}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}} = -\frac{(1+\eta)\bar{\lambda}}{\nu A}\left(\frac{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}}\right) < 0 \quad (172)$$

□

Proof of Proposition 4. **Firms**

$$\frac{d\tilde{N}}{dA} = \frac{\partial \tilde{N}}{\partial A} + \frac{d\tilde{N}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} \quad (173)$$

$$= \frac{(1+\eta)}{\nu\eta A}\tilde{N} - \frac{\tilde{N}}{\bar{\lambda}\eta}\left[\frac{(1+\eta)\bar{\lambda}}{\nu A}\left(\frac{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}}\right)\right] \quad (174)$$

$$= \frac{\partial \tilde{N}}{\partial A}\left[1 - \frac{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}}\right] = \frac{\partial \tilde{N}}{\partial A}\left[\frac{-\frac{d\tilde{C}}{d\bar{\lambda}}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}}\right] > 0 \quad (175)$$

Bonds

$$\frac{d\tilde{B}}{dA} = \frac{\partial \tilde{B}}{\partial A} + \frac{d\tilde{B}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} = \frac{d\tilde{B}}{d\tilde{N}}\frac{\partial \tilde{N}}{\partial A} + \frac{d\tilde{B}}{d\tilde{N}}\frac{d\tilde{N}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} \quad (176)$$

$$= \frac{d\tilde{B}}{d\tilde{N}}\left[\frac{\partial \tilde{N}}{\partial A} + \frac{d\tilde{N}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA}\right] = \frac{d\tilde{B}}{d\tilde{N}}\frac{d\tilde{N}}{dA} \quad (177)$$

From (135) if $N_0 - \tilde{N} \to 0$ then $\frac{d\tilde{B}}{d\tilde{N}} = \frac{\partial \tilde{B}}{\partial \tilde{N}} + \frac{d\tilde{B}}{d\Gamma}\frac{d\Gamma}{d\tilde{N}} = \frac{\tilde{\Omega}}{\Gamma - r}\left(1 + \frac{N_0 - \tilde{N}}{\Gamma - r}\frac{d\Gamma}{d\tilde{N}}\right) \approx \frac{\tilde{\Omega}}{\Gamma - r}$ thus

$$\frac{d\tilde{B}}{dA} \approx \frac{\tilde{\Omega}}{\Gamma - r}\frac{d\tilde{N}}{dA} \gtreqless 0 \implies \operatorname{sgn}\frac{d\tilde{B}}{dA} = \operatorname{sgn} -\tilde{\Omega} \quad (178)$$

Labor:

$$\frac{d\tilde{H}}{dA} = \frac{\partial \tilde{H}}{\partial A} + \frac{d\tilde{H}}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} = \frac{\tilde{H}}{\nu A\eta} + \frac{\tilde{H}}{\nu\bar{\lambda}}\frac{d\bar{\lambda}}{dA} = \frac{\partial \tilde{H}}{\partial A}\left[1 + \frac{\nu A}{\bar{\lambda}}\frac{d\bar{\lambda}}{dA}\right] \quad (179)$$

Substitute out (172)

$$= \frac{\partial \tilde{H}}{\partial A}\left(1 - \frac{(1+\eta)\left(\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}\right)}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\bar{C}}{d\bar{\lambda}}}\right) \tag{180}$$

$$= \frac{\frac{\partial \tilde{H}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\bar{C}}{d\bar{\lambda}}}\left(-\eta\left(\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}}\right) - \frac{d\bar{C}}{d\bar{\lambda}}\right) \tag{181}$$

Substitute out $\frac{d\tilde{H}}{d\bar{\lambda}} = \frac{\tilde{H}}{\bar{\lambda}\eta}$, $\frac{d\tilde{B}}{d\bar{\lambda}} \approx \frac{\tilde{\Omega}}{\Gamma - r}\frac{d\tilde{N}}{d\bar{\lambda}}$ and $\frac{d\bar{C}}{d\bar{\lambda}} = -\frac{1}{\bar{\lambda}^2} = -\frac{\bar{C}}{\bar{\lambda}}$

$$= \frac{\frac{\partial \tilde{H}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\bar{C}}{d\bar{\lambda}}}\frac{1}{\bar{\lambda}}\left(\bar{C} - \tilde{w}\tilde{H} - r\frac{\tilde{\Omega}}{\Gamma - r}\tilde{N}\right) \tag{182}$$

In steady state $\tilde{C} - \tilde{w}\tilde{H} = r\tilde{B}$

$$\frac{d\tilde{H}}{dA} = \frac{\frac{\partial \tilde{H}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\bar{C}}{d\bar{\lambda}}}\frac{1}{\bar{\lambda}}\left(r\tilde{B} - r\frac{\tilde{\Omega}}{\Gamma - r}\tilde{N}\right)$$

From (62) $\tilde{B} - \frac{\tilde{\Omega}}{\Gamma - r}\tilde{N} = B_0 - \frac{\tilde{\Omega}}{\Gamma - r}N_0$

$$\frac{d\tilde{H}}{dA} = \frac{\frac{\partial \tilde{H}}{\partial A}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{d\bar{C}}{d\bar{\lambda}}}\frac{r}{\bar{\lambda}}\left(B_0 - \frac{\tilde{\Omega}}{\Gamma - r}N_0\right)$$

□

Proof of Proposition 5. **Labor**: Totally differentiating $H = H(\bar{\lambda}, N, A)$ keeping N fixed yields.

$$\frac{dH(0)}{dA} = \frac{dH}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA} + \frac{\partial H}{\partial A} \tag{183}$$

$$= -\frac{\partial H}{\partial A}\left[\frac{(1+\eta-\nu)(w\frac{dH}{d\bar{\lambda}} + r\frac{dB}{d\bar{\lambda}}) - \nu\frac{dC}{d\bar{\lambda}})}{\nu\left(w\frac{dH}{d\bar{\lambda}} + r\frac{dB}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}\right)}\right] \tag{184}$$

As in the long-run case, the income and substitution effects of a technological improvement work in opposite directions. The difference between the long-run and impact multiplier is accounted for by the effect of entry, so that

$$\frac{dH(0)}{dA} - \frac{dH(\infty)}{dA} = \frac{dH}{dN}\frac{dN}{dA} = \frac{dH}{dN}\left[\frac{\partial N}{\partial A} + \frac{dN}{d\bar{\lambda}}\frac{d\bar{\lambda}}{dA}\right] \tag{185}$$

$$= \frac{dH}{dN}\frac{\partial \tilde{N}}{\partial A}\left[\frac{-\frac{d\bar{C}}{d\bar{\lambda}}}{\tilde{w}\frac{d\tilde{H}}{d\bar{\lambda}} + r\frac{d\tilde{B}}{d\bar{\lambda}} - \frac{dC}{d\bar{\lambda}}}\right] \tag{186}$$

$$\text{sgn}\left[\frac{dH(\infty)}{dA} - \frac{dH(0)}{dA}\right] = \text{sgn}\ \ H_N = \text{sgn}\left[1-\nu\right]$$

Wages:

$$\frac{dw(0)}{dA} = \frac{1}{\mu}Y_{HH}\frac{dH(0)}{dA} + \frac{w}{A\nu} \tag{187}$$

Hence

$$\frac{dw(0)}{dA} - \frac{dw(\infty)}{dA} = \frac{1}{\mu}Y_{HH}\frac{dH(0)}{dA} \tag{188}$$

$$\text{sgn}\left[\frac{dw(0)}{dA} - \frac{dw(\infty)}{dA}\right] = \text{sgn}\left[\nu - 1\right] \tag{189}$$

The difference between the long-run and short run wage effect depends on whether an increase in employment increases the MPL ($\nu > 1, Y_{HH} > 0$), or decreases it ($\nu < 1$, $Y_{HH} < 0$). □

A.10 Extra Figures

Figure 2 plots a scatter of the Chang and Hong results from Table 2. Red triangles represent the 14 observations that are consistent with our theory.

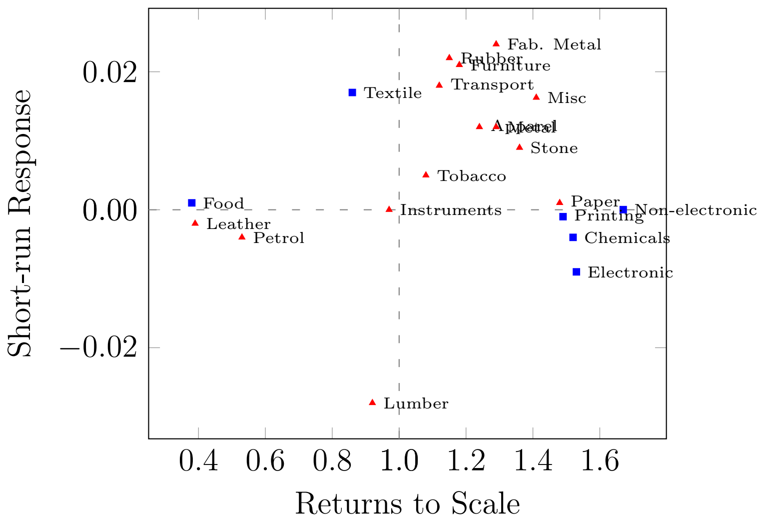

Figure 2: Empirical Evidence