UNEQUAL GAINS, PROLONGED PAIN:
A Model of Protectionist Overshooting and Escalation*

Emily Blanchard† Gerald Willmann‡

July 15, 2020

Abstract

We develop a model of democratic responses to macroeconomic shocks, and show that when economic adjustment is slower than potential political change, economic shocks can trigger populist surges. Applied to trade policy, we show that unexpected changes in world prices or skill biased technological change can induce a surge in economic nationalism and trade protection. Over time, the initial protectionist surge will gradually diminish if and only if educational gains enable less-skilled workers to catch up with the overall economy. The more unequal the initial distribution of the returns to human capital, the greater and longer-lasting the protectionist backlash will be: unequal gains, prolonged pain. Evidence on key data markers suggested by the model exhibits patterns consistent with recent populist support for Brexit and Trump.

JEL Classifications: F5, D7, E6

Keywords: Populism, Protectionism, Overshooting, Dynamic Political Economy, Human Capital, Education, Overlapping Generations, Endogenous Tariffs

*We are grateful for extensive and constructive feedback from our colleagues and seminar participants at the NBER ITI, CEPR ERWIT, RMET, ETSG, EIIT, Villars-sur-Ollon International Trade Conference, SEA, TAMU, Bank of England, Barcelona GSE, Brandeis, Cambridge, CEU, Chicago FRB, Columbia University, Dartmouth, Hitotsubashi, Oxford, LSE, SUFE, Syracuse, UBC, and others.

†Tuck School of Business at Dartmouth and CEPR; emily.blanchard@tuck.dartmouth.edu

‡University of Bielefeld and IfW Kiel; gwillmann@wiwi.uni-bielefeld.de

1 Introduction

Globalization has suffered a spate of sharp democratic rebukes over the past few years, starting with the UK 'Brexit' vote and the US presidential election of trade-skeptic Donald Trump. The subsequent surge in economic nationalism has vexed globalization's cheerleaders, who are quick to point out that despite individual losses for some, the aggregate gains from trade and immigration are positive and, moreover, that technological change is at least as responsible as foreign competition in driving job losses. While these arguments are correct, they fail to appreciate two key forces underpinning today's protectionist groundswell: labor market frictions and rising inequality. This paper incorporates both factors into a workhorse dynamic political economy model to gain new insight on the drivers and consequences of today's rising protectionism.

We show that when economic adjustment is slow and the gains from trade are skewed toward the top, protectionist surges are a natural and long-lasting democratic response to unanticipated macroeconomic changes. Crucially, this prediction holds even when shocks deliver immediate aggregate welfare gains, even if those gains will eventually be shared by a majority of voters, even when the shocks are driven by technology instead of trade, and even in the presence of redistributive income taxes and transfers. The key mechanism underlying these findings is a fundamentally a timing mis-match: structural change takes time, while politics can respond more quickly. So, even if in the long run most individuals would be 'winners' from more open borders, in the short run, many workers suffer when labor market frictions hamper their potential to respond to a changing marketplace. In the immediate aftermath of negative labor market shocks, import-competing workers have a stronger incentive to use tariffs to boost market demand for their labor. Remedial trade protection slows the eventual process of trade adjustment, however, which slows the subsequent rate of political and economic adjustment over time, with long-lasting welfare consequences.

The core of our paper formalizes this insight by developing a dynamic political economy model to identify the short and long-run consequences of labor market frictions in a responsive democratic political environment. We consider unanticipated changes in the terms of trade and skill-biased technological change, and show that the sharp democratic reactions to these macroeconomic shocks may impose long-lasting efficiency costs by distorting future economic decisions. We use the model to evaluate the extent to which domestic economic policies (including income taxes, universal basic income, or education) or multilateral trade agreements will soften or sharpen the political consequences of macroeconomic

shocks for trade policy. A short final section of the paper presents recent data from the US, UK, and comparable trading partners in the context of our theory, and finds patterns consistent with recent populist support for Brexit and Trump.

Our framework provides a markedly different lens through which to understand contemporary trade policy than the well-understood "Protection for Sale" politics emphasized by Grossman and Helpman (1994). While there is substantial evidence that special interest lobbying has played a central role in shaping trade policy for most of modern history,[1] the same explanations are unsatisfying in the context of today's populist protectionism, which is squarely at odds with corporate interests. By emphasizing the role of popular politics in the presence of wage inequality, labor market frictions, and rapid technological and global market changes, our model delivers predictions consistent with recent experience. Our findings also complement important recent research on identity politics:[2] to the extent that rising economic vulnerability exacerbates underlying group allegiances, then trade and technology shocks may trigger rising polarization in political identity, consistent with recent evidence found by Autor, Dorn, Hanson, and Majlesi (2020) in response to the terms-of-trade 'China Shock'.

At the same time, the key insights of this paper apply beyond today's surge in protectionism. We tailor our model to address the recent surge in tariffs, but the key finding is much more general: differential frictions between economic and political change can induce sharp political swings with long-lasting consequences. Both this idea and the general theoretical mechanisms that we highlight are readily applicable to other policy contexts, including climate change, immigration, and tax or entitlement reform. As we demonstrate formally through the specific lens of our model, the interplay between slow economic adjustment and rapid political response can generate rich, non-monotonic transition dynamics. This insight offers a political-economy analog to the seminal Dornbusch (1976) finding, that the marriage of sticky prices with immediate adjustment of market expectations generates non-monotonic exchange rate "overshooting" dynamics.[3]

The model features a small open economy with overlapping generations of heterogenous workers who make endogenous human capital investments. In each generation, young

[1] See Irwin (2017) for the authoritative history of US trade policy, and Gawande and Bandyopadhyay (2000), Goldberg and Maggi (1999), and Matschke and Sherlund (2006) for evidence consistent with the Grossman-Helpman model specifically.

[2] e.g. Gennaioli and Tabellini (2019) and Grossman and Helpman (2020)

[3] We deliberately use the term *overshooting* to evoke and pay tribute to Dornbusch (1976).

workers form rational expectations over the future (exogenous) macroeconomic environment and (endogenous) policy outcomes. We model the policy instrument as a tariff, which generates a clear tradeoff between aggregate welfare and the distribution of income. Policy is determined by majoritarian voting according to a median voter rule, in the tradition of Mayer (1984).[4] We consider permanent exogenous shocks to the terms of trade and skill-biased technology, which we show can have commensurate political consequences. We focus on the empirically relevant scenario in which a macroeconomic shock increases aggregate income, but whose benefits accrue disproportionately to the most skilled/highest income individuals.

The theoretical analysis generates three key insights. First, differences in the potential speed of adjustment between economic and political change can lead to policy volatility. As long as politics can respond to shocks more quickly than labor markets can adjust, then the short run response will be an increase in trade protection – even if the shock will eventually lead to lower tariffs. Moreover, this surge in protectionism will slow the subsequent process of political and economic adjustment by blunting the incentive for younger workers to acquire human capital: although the immediate tariff spike will not fully offset the initial increase in the skill premium, the protectionist surge will reduce the extent to which younger workers shift into the higher skill sectors.

Second, skewness in the distribution of human capital plays a critical role in both the short and long run. At the time of the shock, greater inequality leads to a sharper initial protectionist surge and thus a longer and more costly adjustment process. In the long run, inequality itself is endogenous and there are two possibilities. The optimistic scenario is *protectionist overshooting*, which obtains if education allows adversely-affected workers to 'catch up' to the rest of the economy. If so, the process takes time, but exhibits a virtuous cycle: as education rises, inequality falls, which increases support for trade; lower tariffs then induce further educational investment, and so on. Alternatively, it is entirely possible that the shock will exacerbate the underlying skewness in the distribution of the returns to human capital, even after workers have had time to upgrade their skill sets. Under this scenario, the initial shock will induce a greater increase in human capital

[4] Alesina and Rodrik (1994) argue that the median voter rule acts as a tractable stand-in for nearly any political environment in which the underlying distribution of voters' preferences matters; under more general political systems, a different moment of the population distribution (rather than the median) will drive formal results, but the upshot remains the same: the overall distribution of gains and losses – not just the aggregate – is critical in determining policy.

at the top of the distribution than at the bottom, and the long run equilibrium will be characterized by *protectionist escalation*: after the initial protectionist surge, inequality rises *despite* increasing education levels, and the tariff will continue to rise via an oscillating transition path, converging to a higher steady state level.

Third, we demonstrate that skill-biased technological change can mimic the effects of a terms-of-trade improvement in triggering a protectionist backlash. In our model, both of these shocks drive up the skewness in the returns to human capital, with commensurate political effects. Thus, a populist backlash against globalization could be caused by technology, not trade: even if automation is *exclusively* responsible for today's increasing economic polarization, the political consequences for globalization may be the same. More broadly, anything that increases dispersion in the distribution of earnings can sharpen voters' incentives to tilt market wages in their favor, using trade policy or other means. In a democracy, economic nationalism may be an inevitable and natural consequence.

We use the model to evaluate the extent to which protectionist surges are exacerbated or dampened by other domestic economic policies. We show that progressive income taxes or unconditional redistribution (e.g. universal basic income) will not eliminate protectionism, even if they reduce income inequality: as long as some part of workers' earnings are linked to market wages, voters will have an incentive to manipulate trade in order to boost demand for their own labor. At the same time, progressive taxes and conditional transfers risk discouraging investment in education (as do tariffs on imports of low-skill goods). Education subsidies or reforms are more promising. They can both encourage human capital formation and reduce protectionist pressure, but only if they induce convergence in the distribution of human capital. To the extent that education policies increase human capital disproportionately among those workers already at the top of the distribution, they may only worsen political polarization and, thus, protectionism. In a separate extension, the model highlights the importance of escape clauses in multilateral trade rules. Absent safeguard flexibilities, a short-term protectionist spike could lead to a permanent trade war.

We offer empirical context for our theoretical analysis using data from the US, UK, and other labor markets. Theory guides us to look for evidence of two conditions, which if satisfied would predict protectionist overshooting or escalation in response to recent macroeconomic changes. The first condition is that the returns to human capital, and thus gains from trade, are concentrated at the top. Though by no means a perfect measure, we compare trends in mean and median household (gross) income to proxy the evolution of 'unequal

gains' over time and across countries. The second condition is that labor market adjustment is in fact "sticky". Labor market frictions are notoriously difficult to estimate, especially across countries, but intergenerational earnings elasticities offer a rough indication of the extent to which workers can overcome initial differences and reduce the skewness of income differences over time. Data on both indicators suggest that the US and UK are unusual relative to otherwise comparable OECD countries: economic inequality and intergenerational income immobility are highest where economic nationalism recently won electoral success under Brexit and Trump.

The paper proceeds as follows. The next section reviews the important and diverse related literature that precedes us. Section 3 then presents our model and characterizes economic and political steady states. Section 4 examines the transition dynamics following a large permanent terms-of-trade shock and, in an immediate extension, demonstrates the nearly isomorphic effects of skill-biased technological change. In Section 5, we use the model to shed light on domestic and multilateral policies that may exacerbate or mitigate populist protectionist surges. Section 6 presents data on empirical indicators suggested by the model and Section 7 concludes.

2 Related Literature

This paper is motivated by a series of important recent empirical findings, which together document the prevalence of labor market frictions that can exacerbate inequality in the gains from globalization, the potential for protectionist surges, and the recent rise of political polarization in Western democracies. This large and growing relevant literature includes Artuç, Chaudhuri, and McLaren (2010), Autor, Dorn, and Hanson (2013), and Dix-Caneiro (2014) who, among others, highlight the important role that adjustment costs play in shaping the distributional consequences of trade. Empirical findings by Bown and Crowley (2012) and Hillberry and McCalman (2011) suggest that the use of flexible protectionist policy instruments (anti-dumping cases and other temporary trade barriers) can and do surge temporarily in response to global economic shocks,[5] while Piketty (2018) documents a parallel escalation of inequality, populism, and nativism in the US, UK, and France over the past half-century.

[5]Bown and Crowley (2012) find evidence of sharp protectionist responses to recessionary business cycles, while Hillberry and McCalman show that import surges (consistent with sharp terms-of-trade changes) precipitate protectionist anti-dumping filings in the U.S, which are designed to sunset over time.

Our interest in the trade policy impact of the terms-of-trade "China Shock" is also shared by a recent strand of literature on identity politics. Grossman and Helpman (2020) use a Heckscher-Ohlin setup in which factor owners identify politically with their peers (and possibly with the entire population), and find that changes in political identification can also lead to a protectionist reallignment. Gennaioli and Tabellini (2019) find a similar relationship between political identity and trade preferences in a a public finance model with endogenous redistribution (notably, trade exposure is assumed to be orthogonal to income distribution in their framework). Although these models differ in their precise characterization of political identity,[6] the fundamental insight is the same: trade and technology shocks can drive major shifts in political identification, consistent with the recent evidence documented by Autor, Dorn, Hanson, and Majlesi (2020). We view our contribution as complementary to these findings. We find the potential for similar political dynamics even in the absence of a political identity filter; i.e. with rational, narrowly (neoclassically) self-interested individuals. To the extent that shocks create new economic vulnerabilities that also exacerbate underlying political group-identity allegiances, our findings imply that trade or technology shocks could further exacerbate "culture war" politics *in tandem with* rising polarization over economic policy.[7] At the same time, our more standard neoclassical approach gives us new insight into the conditions under which protection may recede in the longer term and what levers could be used to bring about broader political support for more inclusive globalization.

While our study of endogenous political transition dynamics is motivated by recent data, our model is built on a long tradition of work in trade, political economy, and macroeconomics. In our approach to modeling endogenous trade policy with heterogeneous voters, we follow in the tradition of Mayer (1984), whose seminal model links inequality in the (static) distribution of physical capital with democratic support for trade protection in capital-abundant countries.[8] At the same time, the political hysteresis in our model continues the tradition of Fernandez and Rodrik (1991) and Jain and Mukand (2003), who demonstrate the potential for endogenous resistance to trade reform due to uncertainty. From a modeling perspective, our paper recalls the "putty-clay" labor market structure in

[6] In the first model, an individual's utility depends in part on her group's average welfare, while in the second, political identification influences an individual's subjective beliefs about her income prospects

[7] Political provocateurs have long used identity politics (often, by race) to manipulate political divisions over economic policy. For instance, "makers versus takers" rhetoric has been deliberately weaponized to reduce lower-income voters' support for post-tax redistribution.

[8] See Dutt and Mitra (2002) and Dhingra (2014) for empirical support.

Matsuyama (1992) to generate rich and plausible transition dynamics.

Our work is also reminiscent of Brainard and Verdier (1997), who develop a model in which declining import-competing industries can slow their decline via costly lobbying for protection. In complementary work, Staiger and Tabellini (1987) highlight the importance of time consistency (or its absence) in driving "excessive" protection, which can occur if long-lived governments cannot pre-commit to future free trade. While our overlapping generations framework is quite different from theirs – by definition, the democratically most-preferred tariff is not "excessive" – their broader implication is also salient here: tariff commitments can play an important role in structural change, as we later discuss in the context of multilateral escape clauses. We also build on our previous work in Blanchard and Willmann (2011), to study transition dynamics explicitly.[9]

In the macro literature, Alesina and Rodrik (1994), Persson and Tabellini (1994), Krusell and Ríos-Rull (1996), Bassetto (1999), and Hassler, Rodríguez Mora, Storesletten, and Zilibotti (2003), also feature overlapping generations models with slow adjustment, but none of these models allow for both the differential speed of real versus political adjustment and the endogenous evolution of political preferences (e.g. via income), which together give rise to our overshooting and escalation dynamics. An analysis in a similar setting as ours, but again, without the same transition mechanics, has also been undertaken in the area of migration; see Storesletten (2000) for a seminal contribution and Razin, Sadka, and Suwankiri (2011) for a broader overview.

More recently, Acemoğlu, Naidu, Restrepo, and Robinson (2015) highlight the interplay between democracy and redistribution and find empirical support for the importance of the politically pivotal middle class, particularly in promoting redistribution and structural change through secondary schooling.[10] Outside the political economy framework, but also closely related is the important work by Helpman, Isthoki, and Redding (2010) and Acemoğlu, Gancia, and Zilibotti (2015), who demonstrate the potential for increased openness, offshoring, and endogenous skill-biased technical change to increase inequality through complementary channels.

Finally, our work responds to the forceful call by Acemoğlu and Robinson (2013) to

[9] Our previous work examined the potential for switching between steady states in a setting with a binary policy choice, binary skill acquisition decisions, and multiple equilibria, but adopted modeling restrictions that precluded the study of transition paths.

[10] The paper also raises an important qualification to our median voter approach to the extent that political power is captured completely by richer segments of the population. We return to this issue later in the paper.

recognize the feedback effects between economic reforms and political outcomes. In the process, we also offer a political-economy counterpart to Antras, deGortari, and Itskhoki (2016), who emphasize the importance of accounting for inequality in modern trade models. Their work provides a compelling critique of the standard Kaldor-Hicks criterion for measuring the welfare consequences of changes in trade patterns; our findings further challenge the static Kaldor-Hicks benchmark by identifying an additional potential long run welfare cost of inequality via endogenous political responses to macroeconomic shocks.

3 A Model of Protectionist Overshooting and Escalation

This section presents an overlapping generations model with heterogenous workers, endogenous human capital formation, and democratic trade policy determination. In our small-country open-economy two-commodity model, two-period lived heterogeneous agents decide how much costly education to acquire during the first period of their lives and reap the benefits of acquired human capital in the second period. Trade policy is determined each period through majority voting. The decisive median voter sets trade policy based on her (existing) level of human capital and the terms of trade. Thus, the equilibrium policy outcome in each period is determined by the population's education decisions from the previous period. The central importance of the stock of human capital on current trade policy decisions and the slow adjustment of this structural variable introduce political hysteresis, even in the absence of uncertainty.[11]

We model trade policy as an ad-valorem tariff on imports of goods produced with unskilled labor, and show that starting from a political steady state with a positive, non-prohibitive tariff, an exogenous aggregate terms-of-trade improvement for the country will lead to a protectionist surge: an immediate sharp increase in trade protection. There are then two long-run possibilities. First, if an overall, economy-wide increase in educational investment induces income convergence, equilibrium will be characterized by *protectionist overshooting*: once workers have time to adjust the now-higher global demand for skills, political polarization will gradually abate and the tariff will slowly fall. Alternatively, if an increase in the global skill-premium induces workers at the top of the income distribution to invest in human capital even faster than their counterparts lower on the income ladder,

[11] Uncertainty over future policy outcomes would introduce additional policy hysteresis via the uncertainty-driven status-quo bias mechanism à la Fernandez and Rodrik (1991) or Jain and Mukand (2003); our mechanism obtains despite the absence of uncertainty.

protectionist escalation will ensue: over time, the politically pivotal median voter will be left even further behind by the most skilled workers and the new steady state tariff will be even higher than the initial protectionist surge.

In a brief extension, we show that unanticipated skill-biased technological change (SBTC) is virtually isomorphic to a terms-of-trade shock in generating protectionist dynamics.

3.1 The Economy

Consider a small open home economy that produces, consumes, and trades two goods: a skill-based good, S, which requires skilled labor to produce, and a basic good, U, produced using unskilled labor. Both goods are produced under perfect competition with constant returns to scale technologies. We assume that our small country has comparative advantage in the skill-based good, S, adopting the perspective of an industrialized country. An import tariff applied on imports of the basic good U thus depresses the domestic relative price of the skill-based good. Designating U as numéraire, the domestic relative price of good S is given by $p \equiv \frac{p^w}{\tau}$, where p^w represents the exogenous world relative price of the skill-based good and τ is equal to one under free trade and strictly greater than one under a tariff.[12] Note that our simple production structure limits the price vector to one relative price, while still allowing us to capture Stolper-Samuelson forces in a short-hand way.

The home country is populated by a continuum of heterogeneous agents. Individuals differ in *inherent advantage*, which is fixed at birth and captures initial and immutable differences in characteristics – ability or other accidents of birth (e.g. location, per Chetty, Hendren, and Katz (2016)) – that will ultimately combine with acquired education to realize an individual's human capital. 'Advantage', indexed by a, is assumed to be distributed continuously over the unit interval with cumulative distribution function $F(a)$ and corresponding density function $f(a)$. Agent $a = 0$ is the least advantaged of her generation, and agent $a = 1$ the most advantaged.

Individuals live for two periods; thus at any point in time, two generations, the young (denoted by ^{y}) and the old (denoted by ^{o}), comprise the total population. The population of each generation is normalized to one. We refer to the generation that is young at time

[12] Likewise, $\tau < 1$ represents an import subsidy. Formally, given our choice of numéraire, $\frac{\tau-1}{\tau}$ is the ad-valorem tariff applied to the imported basic good, or equivalently, $t \equiv (\tau - 1)$ is the export tax applied to the domestic price of good S.

t as 'generation t' hereafter. Agents have rational expectations. Finally, we assume that tariff revenue is rebated uniformly across agents within each generation.[13]

Every agent is endowed with one unit of labor in each period of life and is born unskilled. When young, each individual may choose whether to acquire human capital, h, via costly education. Schooling takes time, and so the cost of acquiring human capital is the foregone income from work in the unskilled sector when young. To keep matters simple, we assume that there are no additional pecuniary costs of education, and that education yields no return until the second period of life, when it manifests as human capital. Agents may allocate anywhere from none to all of their per-period (unit) time endowment to schooling. Denoting unskilled labor allocation by l, and duration of education by e, the within-period time constraint is:

$$l + e = 1. \tag{1}$$

Education is an investment: the cost is borne during youth, while the benefits accrue in the future. Thus, in this two-period overlapping generations framework the old have no incentive to acquire additional education in the second period of life. Our structure is thus effectively an extreme case of putty-clay skill 'stickiness' as in Matsuyama (1992).[14]

We assume that every given worker's human capital in the second stage of life is strictly increasing both in her innate advantage, a, and the extent of education she acquired when young, e; that education and inherent advantage are complementary in creating human capital; and that the marginal return to education in terms of human capital is decreasing in education.[15] Defining human capital to be a twice-continuously differentiable function of education and ability, these assumptions can be summarized as follows:

[13]This intra-generational rebating assumption removes any potential intergenerational transfer motivation for tariffs, which both isolates the education-driven distributional motivations that are our focus, and helps eliminate nuisance equilibria (see footnote 21).

[14]More generally, we could assume only that the adjustment cost increases as a worker gets older. What is crucial for our key mechanism and results is simply that economic adjustment is slower than political change: skill stickiness is one of many ways to establish this sort of economic hysteresis in the (human) capital stock.

[15]The complementarity assumption generates the single crossing condition necessary to ensure that higher a workers self-select into higher education levels (assortative matching), while concavity ensures the second order condition for individuals' optimal education decisions is satisfied.

Assumption 1.

$$\frac{\partial h(a,e)}{\partial a} > 0, \qquad \frac{\partial h(a,e)}{\partial e} > 0, \tag{2}$$

$$\frac{\partial^2 h(a,e)}{\partial a \partial e} > 0, \qquad \frac{\partial^2 h(a,e)}{\partial e^2} < 0. \tag{3}$$

Production and Education. The technology for basic good production is deliberately simple: one unit of unskilled labor produces one unit of the basic (numéraire) good for all workers, so that the unskilled wage is normalized to one. Production of the skill-based good depends on human capital times a constant productivity shifter, b, according to:

$$x^s = bh \qquad \text{with} \qquad b \geq 1, \tag{4}$$

where b is used later to study the effect of skill-biased technological change.

Each agent chooses her education level to maximize her lifetime utility. Preferences are identical across individuals and additively separable across time. Let each agent's lifetime utility function be given by:

$$u(d^{u,y}, d^{s,y}) + \beta u(d^{u,o}, d^{s,o}), \tag{5}$$

where $\beta > 0$ represents the inter-temporal discount factor, $d^{g,k}$ denotes the individual's consumption of good $g \in \{s(killed), u(nskilled)\}$ when she is of age $k \in \{y(oung), o(ld)\}$; and intra-temporal utility is Cobb-Douglas, with: $u(d^u, d^s) = (1-\alpha)\ln d^u + \alpha \ln d^s$. Note that these preferences remove any consumption smoothing motives for skill acquisition.[16] Additionally, with homothetic preferences, intra-period indirect utility later may be written in Gorman form: $v(p)I$, where I denotes current nominal income.

Nominal income for a young worker of any type a in generation t is given by her time in the unskilled labor force plus her share of (intra-generational) tariff revenue, R_t^y:

$$I_t^y(a, e_t) = l_t + R_t^y = 1 - e_t + R_t^y, \ \forall a.$$

Earnings in the second period of life are given by an individual's contribution to basic good output (which is the same for all workers by assumption) plus earnings from skilled good

[16] Under constant marginal utility of income, agents' skill acquisition decisions are orthogonal to savings and wealth. Adding credit markets to the model would also silence any consumption smoothing motive in education decisions, but require more modeling apparatus.

production that accrues to acquired human capital, plus tariff revenue stemming from basic goods imports of the old.[17] For the young worker of generation t and type a, income in the second period of life is given by:

$$I^o_{t+1}(a, e_t) = 1 + bh(a, e_t)p_{t+1} + R^o_{t+1}, \ \forall a.$$

Notice that the return to education is increasing (multiplicatively) in human capital, the skill-biased technological change parameter b, and the relative price of the skill-based good.

Given current and expected prices, which determine the opportunity cost of education and the future returns to human capital, every agent a of each generation t chooses her optimal level of education to solve:

$$\max_e \quad v(p_t, I^y_t) + \beta v(p_{t+1}, I^o_{t+1}) \tag{6}$$

Note that a (uniform) tariff revenue rebate will not influence agents' skill acquisition decisions under our assumption of constant marginal utility of income. The optimal education decision is then given by the first order condition:

$$\beta b \frac{\partial h(a, e)}{\partial e} p_{t+1} = \frac{v(p_t)}{v(p_{t+1})}. \tag{7}$$

Using the definition of the domestic price $p_t \equiv \frac{p^w_t}{\tau_t}$ and rearranging yields the optimal education level for each individual as a function of a, current, and future prices – and thus, current and future tariffs and world prices. (Hereafter, we suppress p^w_t and p^w_{t+1} as arguments to economize on notation.)

$$e(a; \tau_t, \tau_{t+1}) \equiv h_e^{-1}\left(a, \left(\frac{v(p_t)}{v(p_{t+1})} \frac{\tau_{t+1}}{\beta p^w_{t+1} b}\right)\right) \text{ where } p_t = \frac{p^w_t}{\tau_t} \forall t \tag{8}$$

and (with a slight abuse of notation), we use $h_e^{-1}(\cdot)$, to indicate the inverse of the first derivative of $h(a, e)$ with respect to e.

Our assumptions over human capital formation, $h(a, e)$, ensure existence and uniqueness of the optimal education function, $e(a; \tau_t, \tau_{t+1})$.[18] Moreover,

Lemma 1. *The optimal education choice, $e(a; \tau_t, \tau_{t+1})$, is strictly increasing in the agent's initial advantage level a, the discount factor, β, and the current and (expected) future domestic relative price of the skill-based good, p_t and p_{t+1}. [Proof in Appendix A.1]*

[17] Results would be qualitatively similar under proportional tariff revenue redistribution per Mayer (1984).

[18] Specifically, the strict monotonicity and concavity of $h(a, e)$ in e guarantees both the invertibility of h_e with respect to e (existence), and strict inequality for the second order condition of (6) (uniqueness).

The following corollary follows immediately, since the tariff is applied to the basic (unskill-intensive) good:

Corollary 1.1. *The optimal education choice, $e(a;\tau_t,\tau_{t+1})$, is decreasing in the current and (expected) future tariff, for all a.*

An agent's optimal education level increases with her inherent advantage due to the complementarity between education and a. For every individual, education is increasing with the relative weight she places on her future (β) and the greater the domestic relative price of the skill-based good when she is young, since both lower the opportunity cost of education relative to the gains. Likewise, a higher relative price of the skill-based good in the future increases the return to education directly.

Recall that young agents provide unskilled labor only when not in school, while all older agents are assumed to produce one unit of unskilled output in addition to any skilled-good output derived from acquired human capital. Aggregating across all agents of both generations at a given time t then yields the output of each good, $\bar{x}_t^s$ and $\bar{x}_t^u$.[19] (Recall that young agents provide unskilled labor only when not in school, while all older agents are assumed to produce one unit of unskilled output in addition to any skilled-good output derived from acquired human capital.) The following summarizes the equilibrium outcome of the model developed so far, taking tariffs and world prices as exogenous.

Definition 1. *Given a sequence of world prices and tariff pairs, (p_t^w,τ_t) $\forall t\in\mathbb{N}$ an* ***economic equilibrium*** *is a list of education decisions by every agent $a\in[0,1]$:*

$$e_t(a)=e(a;\tau_t,\tau_{t+1})=h_e^{-1}\left(a,\left(\frac{v(p_t)}{v(p_{t+1})}\frac{\tau_{t+1}}{\beta p_{t+1}^w b}\right)\right)\ \textit{where}\ p_t=\frac{p_t^w}{\tau_t}\forall t \tag{9}$$

and associated total quantities of each good produced in the country:

$$\bar{x}_t^u = \bar{x}^u(\tau_t,\tau_{t+1})=\left(1-\int_0^1 e(a;\tau_t,\tau_{t+1})f(a)da\right)+1 \quad \forall t \tag{10}$$

$$\bar{x}_t^s = \bar{x}^s(\tau_{t-1},\tau_t)=b\int_0^1 h(a,e_{t-1}(a;\tau_{t-1},\tau_t))f(a)da. \quad \forall t \tag{11}$$

for every period t in time.

Notice that unskilled output depends on current and future tariffs and prices, via the young cohort's education choices, whereas skilled output depends on past and current prices via the older generation's previous education decisions.

[19] Note that each generation is normalized to mass one, the aggregates are thus per capita averages, which explains the notation.

An *economic steady state* is then simply an economic equilibrium that obtains under a constant world price, p^w and a constant tariff τ such that the domestic price $p = \frac{p^w}{\tau}$ is also constant. In what follows, use overscript tilde ($\tilde{}$) to denote the steady state values of endogenous variables. Steady state functions are defined using a single tariff argument without time subscripts; i.e. $e(a;\tau) \equiv e(a;\tau_t,\tau_{t+1})$ where $\tau_t = \tau_{t+1} = \tau$.

Definition 2. *Given a constant world price p^w and tariff τ, an* ***economic steady state*** *is a list of constant education decisions:*

$$\tilde{e}(a) = e(a;\tau) = h_e^{-1}\left(a, \left(\frac{\tau}{\beta p^w b}\right)\right)\Bigg), \text{ where } p = \frac{p^w}{\tau}, \forall a \in [0,1] \tag{12}$$

and constant associated output quantities:

$$\tilde{x}^u = \bar{x}^u(\tau) = \left(1 - \int_0^1 \tilde{e}(a)f(a)da\right) + 1 \tag{13}$$

$$\tilde{x}^s = \bar{x}^s(\tau) = b\int_0^1 h(a,\tilde{e}(a))f(a)da \tag{14}$$

that obtain at every period t in time.

Consumption is determined in turn by prices and income, while imports and exports are the difference between domestic production and consumption. For a small open economy, aggregate national income is maximized under free trade, which corresponds to equations (12) through (14) evaluated at $\tau = 1$.

3.2 The Political Process

We model the political process as a direct democracy over trade policy, in which only the old generation holds suffrage rights.[20,21] At the beginning of each period, voters choose the current period trade policy, which subsequently determines the relative price and thereby

[20] By limiting voting to the old, we are able to rule out a host of nuisance equilibria that otherwise arise via self-fulfilling expectations. As Hassler, Storesletten, and Zilibotti (2007) point out, limiting voting to the old generation is observationally equivalent to the assumption that elections are held at the end of each period, at which point policy is set for the subsequent period; the old are then assumed to abstain because they will not live to experience the consequences.

[21] See Blanchard and Willmann (2011) and Hassler, Storesletten, and Zilibotti (2003) for models in which both the young and old generations vote. In the first, a binary referendum framework keeps the model tractable at the expense of transition dynamics; in the second, the young side universally with the old poor in taxing the old rich, which again ensures tractability.

the real return to human capital for that period. The vote each period takes place before young agents decide on skill acquisition and before production and consumption occurs. The diagram below illustrates the within-period sequencing.

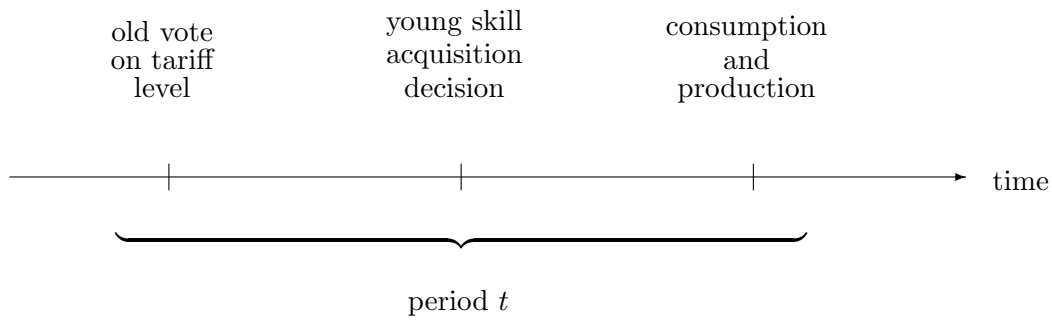


Figure 1: Within-period Sequencing.

The tariff preferences of the electorate are defined as follows. At time t, we denote the distribution of the (given, by then) education levels among the currently-old cohort using $\mathbf{e}_{t-1}$, and use $e_{t-1}(a)$ to represent the education of each individual (again, fixed at time t). From here, each old agent's most preferred trade policy is defined implicitly by:

$$\begin{aligned}\tau_t(a;\mathbf{e}_{t-1}) = \quad & \arg\max_{\tau_t} \quad V^o\big(p_t, I_t^o(a,\mathbf{e_{t-1}})\big) \quad \text{where} \\ & I_t^o(a,\mathbf{e_{t-1}}) = 1 + bh(a,e_{t-1}(a))p_t + R_t^o(\tau_t,\mathbf{e}_{t-1}), \text{ and} \\ & R_t^o(\tau_t,\mathbf{e}_{t-1}) = \frac{\tau_t - 1}{\tau_t}\bar{M}_t^{o,u}(\tau_t,\mathbf{e}_{t-1}) = (\tau_t - 1)p_t\bar{E}_t^{o,s}(\tau_t,\mathbf{e}_{t-1}).\end{aligned}$$

where $\bar{M}_t^{o,u}(\tau_t,\mathbf{e}_{t-1})$ and $\bar{E}_t^{o,s}(\cdot))$ denote per-capita imports of good U and exports of good S among the old generation at time t.[22] Using Roy's identity, the first order condition of the maximization problem can be written as:

$$V_\tau^o = v_I\left(\underbrace{[x_t^{o,s}(a;\mathbf{e}_{t-1}) - d_t^{o,s}(a;\tau_t,\mathbf{e}_{t-1})}_{\equiv E_t^{o,s}(a;\tau_t,\mathbf{e}_{t-1})} - \bar{E}_t^{o,s}(\tau_t,\mathbf{e}_{t-1})]\frac{\partial p_t}{\partial \tau_t} + (\tau_t - 1)p_t\frac{d\bar{E}_t^{o,s}}{d\tau_t}\right) = 0,$$

where $E_t^{o,s}(a)$ denotes the *individual net export position* (the difference between a given worker's production and consumption of good x) of an old individual of type a. Rewriting

[22] $\bar{E}_t^{o,s} \equiv \int_a E_t^o(a;\tau_t,\mathbf{e}_{t-1})f(a)da = \int_a [x_t^{o,s}(a;\mathbf{e}_{t-1}) - d_t^{o,s}(a;\tau_t,\mathbf{e}_{t-1})]f(a)da$ where $d_t^{o,s}(a;\tau_t,\mathbf{e}_{t-1}) = (\alpha/p_t)I_t^o(a;\tau_t,\mathbf{e_{t-1}}))$ is individual a's consumption of good S.

again yields:

$$V_\tau^o = v_I \frac{p_t}{\tau_t}\left(\underbrace{-\Delta(a;\mathbf{e}_{t-1})}+(\tau_t-1)\tau_t \underbrace{\frac{d\bar{E}_t^{o,s}}{d\tau_t}}_{(-)}\right) = 0, \tag{15}$$

where we define the net-skill position of a voter of type a:

$$\Delta(a;\mathbf{e}_{t-1}) \equiv (1-\alpha)b\left[h(a,e_{t-1}(a)) - \underbrace{\int_a h(a,e_{t-1}(a))f(a)da}_{\equiv \bar{h}(\mathbf{e_{t-1}})}\right]. \tag{16}$$

The $\Delta(\cdot)$ denotes the relative net export position of an (old) individual a, *relative* to the average export position within her generation. Since this term plays a central role in the remaining analysis, it is worth pointing out two important properties. First, $\Delta(a;\mathbf{e_{t-1}})$ is fixed at the beginning of time t, *before* voting occurs, as human capital investments were decided by the old when they were young. Second, notice that $\Delta(a;\mathbf{e_{t-1}})$ amounts to $(1-\alpha)$ times the individual's bias in human capital *relative* to the average level of human capital in her generation, $\bar{h}(\mathbf{e_{t-1}}) \equiv \int_a h(a,e_{t-1}(a))f(a)da$. This is because everyone consumes a share α of income of the skilled good under Cobb-Douglas preferences.

The role of individual level heterogeneity in shaping tariff preferences is immediately clear from equation (15). The relative net export position $\Delta(\cdot)$ captures an individual's self-interested motive to use the tariff to distort the wage distribution in her favor, while the second term represents the familiar aggregate efficiency cost of trade restrictions, which is borne by all individuals, and which is minimized by choosing free trade. Starting from free trade, the marginal efficiency cost of changing the tariff is vanishingly small, but the distributional consequences are not. Thus, any relatively unskilled individual for whom $\Delta(a;\mathbf{e}_{t-1}) < 0$ will prefer a strictly positive tariff. Conversely, higher a agents whose net skill position is above the mean ($\Delta(a;\mathbf{e}_{t-1}) > 0$) prefer to subsidize trade. It is only a razor's edge *average* agent, $\hat{a}$, whose individual net skill position perfectly mirrors the mean of her entire generation — that is, for whom $\Delta(\hat{a};\mathbf{e}_{t-1}) = 0$ — who will vote for free trade.[23]

These individual policy preferences reflect the same underlying intuition as the "political cost-benefit ratio" in Rodrik (1994). Starting from free trade, the marginal benefit of using the tariff to redistribute income is strictly positive for any individual who is not herself a perfect mirror of the economy overall. The greater the difference between a voter's

[23]Note that $\hat{a}$ need not coincide with the first moment of a, as the mapping from a to h will in general not be linear.

own net-skill position relative to her generation, the greater her motive to use tariffs to tilt the wage distribution in her favor, even at the expense of overall efficiency.

We summarize the properties of trade policy preferences as follows:

Lemma 2. *The preferred tariff of an old individual a at time t, $\tau(a, \mathbf{e}_{t-1})$, is strictly positive (negative) iff $\Delta(a, \mathbf{e}_{t-1}) < 0 (> 0)$. Moreover:*

$$\frac{\partial \tau(a, \mathbf{e}_{t-1})}{\partial a} < 0, \tag{17}$$

$$\frac{\partial \tau(a, \mathbf{e}_{t-1})}{\partial e_{t-1}(a)} < 0, \tag{18}$$

$$\frac{d\tau(a, \mathbf{e}_{t-1})}{da} < 0. \tag{19}$$

Redefining the function as $\tau(\Delta(a, \mathbf{e}_{t-1}), \bar{E}_t^{o,s}(\mathbf{e}_{t-1}))$, the most preferred tariff is strictly decreasing in Δ:

$$\frac{\partial \tau(\Delta, \bar{E}_t^{o,s})}{\partial \Delta} < 0. \tag{20}$$

[Proof in Appendix A.2]

Lemma 2 formalizes the earlier intuition that more educated voters, and those with greater initial advantages prefer freer trade. All voters with a below-average level of human capital (whether due to accidents of birth (lower a), limited education, or both) prefer strictly positive tariffs, which would tilt real wages in their favor. Individuals with above-average human capital prefer negative tariffs (equivalently,[24] export subsidies), which would further magnify the returns to human capital, tilting real wages in their direction.

Voting. Trade policy is determined by majority vote. Every agent votes for her most preferred tariff policy, $\tau \in (0, \tau^P]$, where τ^P denotes the prohibitive tariff level (and hence a return to autarky) and any $\tau < 1$ indicates an import subsidy. Under the monotonic tariff preferences described in Lemma 2, the median voter, denoted a^m, is decisive. We restrict attention to sincere (and implicitly compulsory) voting to rule out nuisance equilibria.[25] We also abstract from bureaucratic or time costs of changing tariff regimes.

Political equilibrium is composed of two parts: the sequence of tariffs over time as a function of education, and the sequence of education decisions as a function of tariffs. As shown before, equilibrium education is determined by current and expected prices under

[24] by Lerner Symmetry

[25] See Mayer (1984) for a formal treatment of voting costs and probabilistic voting in the median voter environment.

rational expectations according to (8). The equilibrium tariff sequence can be summarized by a trade policy rule that describes the mapping from the state of the world to the then-old median voter's most preferred tariff policy. This trade policy rule has two key features. First, because the median voter is old at the time of the vote, and her welfare does not depend on the decisions of the younger generation, the trade policy rule every period t is independent of future trade policy.[26] Second, since the old median voter's preferred trade policy is determined by the already-fixed distribution of education among her generation, $\mathbf{e_{t-1}}$ serves as the relevant state variable at time t.[27]

Letting $\Delta_t^m \equiv \Delta(a^m; \mathbf{e_{t-1}})$ denote the realized equilibrium relative net export position of the median voter, we define the political equilibrium as follows:

Definition 3. *Given a world price sequence $(p_t^w)_{t\in\mathbb{N}}$, a rational expectations* ***political equilibrium*** *is a sequence of $(\mathbf{e_{t-1}}, \Delta_t^m, \tau_t)$ triples such that starting from $\mathbf{e_0}$ the following holds for all $t \in \mathbb{N} = \{1, ...\infty\}$:*

1. $e_{t-1}(a) \equiv e(a; \tau_{t-1}, \tau_t) = h_e^{-1}\left(a, \left(\frac{v(p_{t-1})}{v(p_t)}\frac{\tau_t}{\beta p^w b}\right)\right) \ \forall a,$

2. $\Delta_t^m \equiv \Delta(a^m, \mathbf{e}_{t-1}) = (1-\alpha)b[h(a^m, e_{t-1}(a^m)) - \bar{h}(\mathbf{e}_{t-1})]$, *and*

3. $\tau_t = \arg\max_{\tau_t} \quad V_t^o(a^m; \mathbf{e_{t-1}}, \tau_t).$

where $p_t = p_t^w/\tau_t \forall t$; $V^o(\cdot)$ denotes the indirect utility for an (old) voter at time t; and $\bar{h}(\mathbf{e}_{t-1}) \equiv \int_a h(a, e_{t-1}(a))f(a)da$ is average human capital among generation $t-1$.

The first condition requires that among the voting-age population at time t, all individuals' skill acquisition decisions are optimal under rational expectations of tariffs over their lifetimes. The second condition defines the net export position of the median voter at

[26] This feature is ensured by the small open economy assumption, and intra-generational tariff revenue rebates, which together imply that the younger generation's education decisions (which do depend on future prices) are immaterial to older voters. Notice that because the optimal tariff rule is independent of future expectations, we do not need to restrict attention to Markov Perfect equilibria, as is customary in many similar models; nuisance equilibria are already ruled out by the model's structure.

[27] Note that the full education distribution $\mathbf{e_{t-1}}$ is actually a strict superset of the relevant state variable at time at time t, since the realized tariff at time t depends only on the median voter's level of human capital and the first moment of the distribution of human capital $\bar{h}(\mathbf{e}_t - 1) \equiv \int_a h(a, e_{t-1}(a))f(a)da$, which enters both $\bar{E}_{t-1}^{o,s}$ and enters $\Delta_{t-1} = (1-\alpha)b[h(a, e_{t-1}(a)) - \int_a h(a, e_{t-1}(a))f(a)da]$.

time t, which depends on the distribution of education among the voting-age population. The third condition requires that the equilibrium realized tariff maximizes the indirect utility of the (older) median voter in period, t. Equilibrium is defined as any sequence of triples $(\mathbf{e_{t-1}}, \Delta_t^m, \tau_t)$ that satisfy these conditions.

We can now define a *political steady state* as an economic steady state in which the status quo trade policy is perpetuated under the existing political process.

Definition 4. *A **political steady state**, summarized by* $(\tilde{\mathbf{e}}, \tilde{\Delta}^m, \tilde{\tau})$ *is characterized by equations* (12) – (14) *and a sequence of constant tariffs* $(\tau_t = \tilde{\tau})_{t\in\mathbb{N}}$ *that jointly satisfy Definition 3:*

$$\tilde{e}(a) = h_e^{-1}\left(a, \left(\frac{\tilde{\tau}}{\beta p^w b}\right)\right) \forall a, \tag{21}$$

$$\tilde{\Delta}^m \equiv \Delta(a^m, \tilde{\mathbf{e}}) = (1-\alpha)b[h(a^m, \tilde{e}(a^m)) - \bar{h}(\tilde{\mathbf{e}})] \tag{22}$$

$$\tilde{\tau} = \arg\max_{\tau} \quad V^o(a^m; \tilde{\mathbf{e}}, \tau). \tag{23}$$

where $\bar{h}(\tilde{\mathbf{e}}) \equiv \int_a h(a, \tilde{e}(a)) f(a) da$.

To save on notation, we refer hereafter to equilibrium and steady state *pairs*, (Δ_t^m, τ_t) and $(\tilde{\Delta}^m, \tilde{\tau})$, which subsume the full distribution of education decisions implied by the underlying model, according to Definition 3.

Steady State Properties. A unique interior steady state exists if there is one (and only one) fixed point solution to equations (21)-(22) such that $\tilde{\tau} \leq \tau^P$ (i.e., up to the non-prohibitive tariff). For the remainder of this paper, we focus on scenarios in which the distribution of the returns to human capital would be skewed toward the top even under free trade,[28] so that the steady state relative net-export position of the median voter is negative, $\tilde{\Delta}^m < 0$, and therefore (by Lemma 2), the steady state tariff is positive, $\tilde{\tau} > 0$.

The following (sufficient) conditions guarantee a unique, stable, interior political steady state:

Assumption 2. *Sufficient conditions for a unique, stable, interior steady state:*

$$\begin{gathered}\left(\tfrac{h_e^2}{|h_{ee}|}\big|_{a^m} - \int_a \tfrac{h_e^2}{|h_{ee}|} f(a)da\right) < \tfrac{\tau^2}{p^w}\left(\bar{h} + \tfrac{1}{bp^w}\right)\alpha((1-\alpha)\tau + 1)), \\ \Delta^m(\tau)\big|_{\tau=1} < 0 \text{ and } \Delta^m(\tau)\big|_{\tau^P} > \tau^{-1}(\Delta^m)\big|_{\tau^P},\end{gathered} \tag{24}$$

where time subscripts are suppressed in steady state.

[28] Empirical wealth and income distributions suggest that this is indeed the relevant case.

Lemma 3. *Under Assumption 2, political equilibrium is unique and stable, both in and out of a steady state. [Proof in Appendix A.3]*

The first condition in Assumption 2 requires that the median voter's most preferred tariff is not overly responsive to small changes in the distribution of human capital – or commensurately, that small changes in the tariff will not generate drastic changes in the relative distribution of human capital. The second set of conditions imply interiority.

Figures 2-3 offer a graphical explanation, by representing steady state first in terms of the median voter's education, $e^m \equiv e(a^m)$ and the steady state tariff, and then in terms of Δ and the steady state tariff. The graph in (e^m, τ) space is more intuitive, while the two panels in (Δ, τ) space align closely with the proofs and prove particularly useful for describing transition dynamics later in the paper.

Figure 2 represents steady state as the intersection of two loci: the median voter's education level, $e^m(\tau; p^w) \equiv e(a^m, \tau; p^w)$, which depends on the (constant) tariff and the world price, and the median voter's most preferred tariff, $\tau(e^m; p^w, \bar{h})$ which depends on the median voter's level of education, holding the aggregate level of human capital ($\bar{h}$) fixed.[29] Both functions are (unambiguously) downward sloping in (e^m, τ) space: the median voter's steady state education schedule is decreasing in the tariff by Lemma 1, and the steady state equilibrium tariff is decreasing in the median voter's steady state education level (again, keeping the mean level of human capital fixed) by Lemma 2. Assumption 2 ensures that the education locus crosses the tariff locus only once and from above. We label the education level at which the median voter would prefer free trade by the benchmark $\hat{e}$, which by definition corresponds to free trade ($\tau = 1$) on the tariff locus. The steady state equilibrium is pinned down by the intersection of the education and tariff loci and labeled $(\tilde{e}^m, \tilde{\tau})$.

Figure 3 offers an alternative depiction of steady state in (Δ^m, τ) space. In each of the panels, equilibrium is again described by the intersection of two loci: the median voter's most preferred tariff as a function of her net skill position, $\tau = \tau(\Delta^m, p^w)$, and the median voter's net skill position as a function of the tariff, $\Delta^m(\tau; p^w)$. In both of these panels, as in Figure 2, the tariff function is strictly downward sloping according to Lemma 2: the less

[29] Moving along this locus is essentially asking "how would the median voter's most preferred tariff change if she, but only she, changed her education level." In equilibrium, of course, $\bar{h}$ depends on the tariff level, and so in steady state, it must be the case that $\tilde{e}^m = e^m(\tilde{\tau}; p^w, \bar{h}(\tilde{e}))$. Figure 3 incorporates this requirement explicitly, and is therefore our preferred illustration.

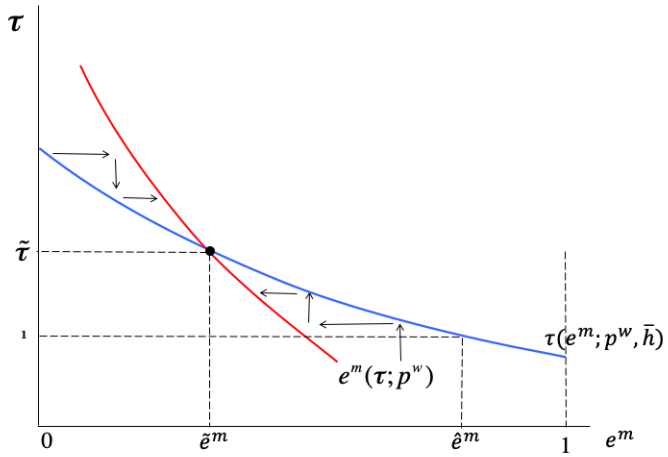


Figure 2: Steady State in (e^m, τ) Space

skilled the median voter is relative to the average of her generation, the more protectionist she will be. As above, if the median voter were perfectly representative of her generation, she would favor free trade, so that $\Delta = 0$ corresponds to $\tau = 1$ on the tariff locus.

The $\Delta^m(\tau; p^w)$ locus is more complex. While individuals' human capital is unambiguously increasing in the domestic relative price of the skilled good (and therefore decreasing in the tariff), Δ^m depends on the *difference* in the median voter's human capital level relative to the rest of her generation. There are two possibilities, both of which are economically interesting and plausible, and both of are represented below.

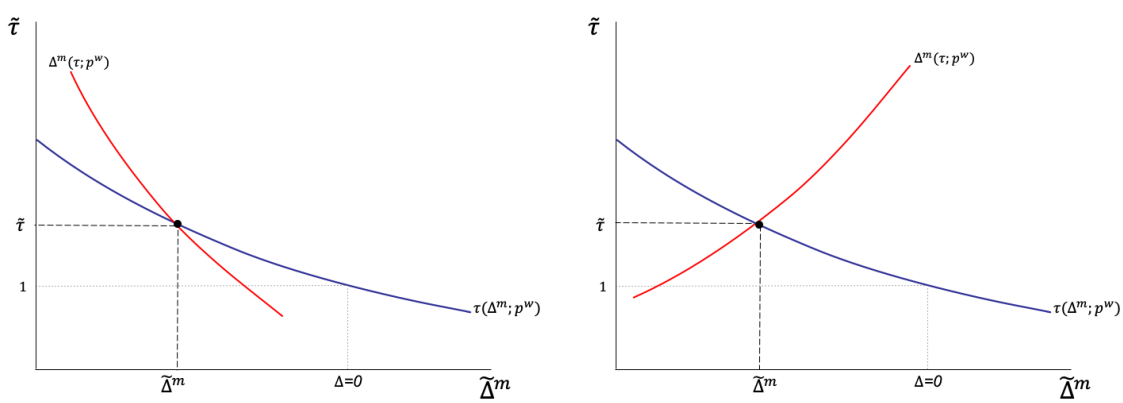


Figure 3: Steady State in (Δ^m, τ) Space. [LHS: $\frac{d\Delta^m}{dp} > 0$, RHS: $\frac{d\Delta^m}{dp} \leq 0$]

The left-side panel of Figure 3 depicts the case in which $\frac{d\Delta^m}{dp} > 0$, which implies that the $\Delta^m(\tau; p^w)$ schedule is downward sloping, as shown. In this scenario, any increase in the domestic skill premium (captured by p in the model) would induce the median voter's human capital level to catch up with the average level of human capital for her generation as a whole. (That is, while everyone would weakly increase her education level in response to an increase in p, the resulting increase in human capital would be faster than average for the median voter.) Thus, lower tariffs would reduce polarization in the distribution of human capital, all else equal. The right-side panel of Figure 3 depicts the opposite case in which $\frac{d\Delta^m}{dp} \leq 0$, which implies that $\Delta^m(\tau; p^w)$ is upward sloping. This case represents the possibility that an increase in the skill premium would cause the median voter to fall further behind the average of her generation. (Despite her increased education in response to an increase in p, the median would fall further behind if the human capital gains from the educational advancements of the rest of her generation outstrip her own gains.) In this case, greater protection would reduce polarization in the distribution of human capital, all else equal.

Our stability condition in Assumption 2 allows for both of the scenarios depicted above, and ensures stability and uniqueness in both cases. Specifically, equation (24) can be rewritten:

$$\frac{d\Delta^m(p)}{dp} \quad < \quad \underbrace{\frac{\tau^2}{p^w} \frac{V_{\tau\tau}}{V_{\tau\Delta^m}}\bigg|_{\tau(\Delta^m)}}_{>0} \forall \Delta^m, \tag{25}$$

which, importantly, allows for $\frac{d\Delta^m}{dp}$ to be positive or negative. Apriori, there is no reason to rule out either case, and indeed, both are economically interesting and plausible.

The conditions for whether $\frac{d\Delta^m}{dp} \lesseqgtr 0$ ultimately depend on the shape of the human capital function, $h(a, e)$ and the underlying distribution of initial advantage, $f(a)$.[30] Intuitively, if the marginal return to human capital is sufficiently low for very high levels of education (for instance, if $h(a, e)$ is sufficiently concave in e, particularly for high-a individuals), then the median voter will be able to catch up when the skill premium rises, so that $\frac{d\Delta^m}{dp} > 0$, consistent with the left-side panel of Figure 3. Conversely, if the marginal return to education remains quite low for the median voter compared to the the average, a higher

[30] From (7) and (16) evaluated for constant p:

$$\frac{d\Delta^m(p)}{dp} > 0 \iff \frac{dh(a^m, e(a^m, p))}{dp} > \int_a \frac{dh(a, e(a, p))}{dp} f(a)da \iff \frac{h_e^2}{|h_{ee}|}\bigg|_{a^m} > \int_a \frac{h_e^2}{|h_{ee}|} f(a)da.$$

skill premium could cause the median voter to fall father behind, $\frac{d\Delta^m}{dp} \leq 0$. We explore both scenarios in detail below.

4 Policy Response to Exogenous Shocks

We now examine the short and long run consequences of a sharp, unexpected,[31] permanent increase in the world relative price of the skilled good, p^w. We adopt the perspective of a relatively skill-abundant, industrialized country in which the initial steady state distribution of human capital is assumed to be skewed toward the top (i.e. the relative net export position of the median voter is negative). This scenario is designed to reflect the circumstances of the "China Shock" – a sharp decline in the world relative prices of goods produced with low-skilled labor – from the perspective of a developed economy like the US or the EU. In an extension, we show that a skill-augmenting technology shock is virtually isomorphic in its political consequences.

4.1 Permanent Increase in the Terms-of-Trade

We use superscript ^{0} for initial steady state values and ^{1} for the new steady state. Starting from an initial political steady state summarized by $(\tilde{\Delta}^{m0}, \tilde{\tau}^0; p^{w0})$ where $\tilde{\Delta}^{m0} < 0$, consider an unanticipated permanent jump in the world price to $p^{w1} > p^{w0}$ at time $t = T$.

As formalized below, the increase in the relative world price of the skilled good will change both the incentives to acquire education and also the preferences over trade policy. We evaluate the consequences of the shock and subsequent adjustment in two stages. First, we describe the properties of the new steady state and then we trace out the transition path by which this new steady state is reached. Throughout, we maintain the regularity conditions in Assumption 2, which ensure equilibrium uniqueness and stability.

4.1.1 The New Steady State

We begin by showing that for any initial distribution of education of the currently-old generation, an increase in the terms of trade (p^w) will further polarize voters' tariff preferences:

[31] With additional modeling apparatus, we can explicitly allow the stochastic shock to be anticipated, i.e. agents rationally expect the shock to happen with a given, low probability as in Baldwin and Robert-Nicoud (2007). This would not change our results qualitatively, as the realization of a shock would still contrast with its expected value. We have therefore chosen to forgo the added complexity.

an initially protectionist voter will become more protectionist, while an initially pro-trade voter will favor even lower (more negative) tariffs.

Intuitively, voters choose trade policy to balance their individual incentive to tilt the domestic relative price in their favor against the shared distortionary cost of trade restrictions. In the initial steady state, these two forces are exactly equal for the median voter, who has chosen the initial steady state tariff, $\tilde{\tau}^0$, to just balance her self-interested motive against the distortionary cost of a marginal tariff change. When the world price changes, this balance is disrupted. Holding the current tariff fixed, an increase in the world relative price would strictly decrease the distortionary cost of increasing the tariff relative to the redistributive motive. Thus, a relatively less skilled median voter would prefer to increase the tariff at least a little bit in response to the increase in p^w. (The opposite would be true if the median voter's were more skilled than average; i.e. if $\Delta^m > 0$.)

The same logic establishes that even the most protectionist median voter would stop short of fully offsetting the terms-of-trade change with the tariff increase. If the median voter were to hold the domestic price fixed by implementing a fully-offsetting tariff, the distortionary cost of the tariff would be strictly higher than at the initial steady state while the redistributive motive would stay the same. Thus, following an increase in p^w, the domestic price p will rise, even if the tariff also increases.

Note that these results hold both in and out of steady state, since the tariff preferences of the old generation depend only on the current world prices and the distribution of education within the older generation. Formally:

Lemma 4. *Polarization effect of an increase in p^w. For any $p^{w1} > p^{w0}$ and any $\Delta_t^m < 0$ ($\Delta_t^m > 0$):*

1. $\tau(\Delta_t^m; p^{w1}) > \tau(\Delta_t^m; p^{w0})$ *($\tau(\Delta_t^m; p^{w1}) < \tau(\Delta_t^m; p^{w0})$), and*
2. $\tau(\Delta_t^m; p^{w1}) < \tau^{FC}$ *($\tau(\Delta_t^m; p^{w1}) > \tau^{FC}$),*

where $\tau^{FC} \equiv \frac{p^{w1}}{p^{w0}}\tau(\Delta_t^m; p^{w0})$ is the fully compensating tariff that would exactly offset the terms-of-trade change, leaving the domestic price unchanged. [Proof in Appendix A.4]

The preceding lemma puts bounds on the new steady state as follows:

Proposition 1. *Steady State response to an increase in p^w. Compared to an initial steady state summarized by $(\tilde{\Delta}^{m0}, \tilde{\tau}^0; p^{w0})$ where $\tilde{\Delta}^{m0} < 0$ and $\tilde{\tau}^0 < \tau^P$, the new steady state under a higher world price $p^{w1} > p^{w0}$, $(\tilde{\Delta}^{m1}, \tilde{\tau}^1; p^{w1})$ has the following properties:*

1. *The new steady state tariff will be less than fully compensating:* $\tilde{\tau}^1 < \tau^{FC} \equiv \frac{p^{w1}}{p^{w0}}\tilde{\tau}^0$, *resulting in a strictly higher domestic price:* $\tilde{p}^1 > \tilde{p}^0$.

2. *The new steady state level of education will be above the old steady state education level for every individual:* $\tilde{e}^1(a) \geq \tilde{e}^0(a) \forall a$.

[Proof in Appendix B.1]

Figure 4 illustrates through the lens of the median voter's education level. Starting from an initial steady state at $(\tilde{e}^{m0}, \tilde{\tau}^0)$, an increase in p^w will cause the steady state education locus $e^m(\tau)$ to shift rightward for all values of τ: intuitively, for any given tariff, an increase in p^w will increase the skill premium and thus the return to education. At the same time, Lemma 4 implies that the new steady state tariff locus will pivot clockwise reflecting the increased dispersion of trade policy preferences among the electorate. The more responsive the education locus to the terms-of-trade shock, the lower the new steady state tariff. Conversely, greater sensitivity of the tariff locus will result in a higher new steady state tariff. Proposition 1 allows us to put additional boundaries on possible relative shifts in the two steady state loci, and implies that the new steady state must lie somewhere in the shaded region. While we know that $\tilde{e}^{m1} \geq \tilde{e}^{m0}$, the new steady state tariff $\tilde{\tau}^1 < \tau^{FC}$ may be above or below the initial tariff, $\tilde{\tau}^0$ depending on where new steady state loci intersect.

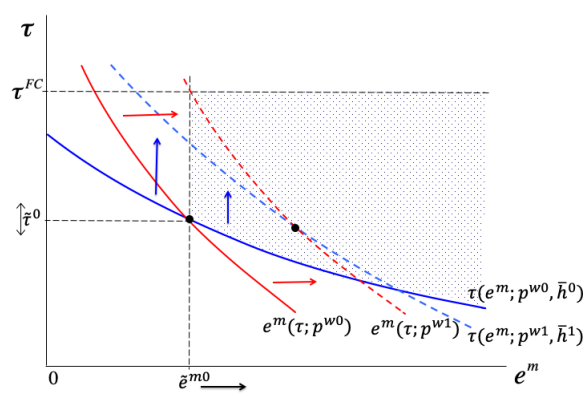


Figure 4: Steady State Response to $p^w \uparrow$ in (e^m, τ)

Figure 5 illustrates depiction in (Δ^m, τ) space. As in the previous figure, the new steady state tariff locus is strictly steeper than the original, pivoting around the free-trade

benchmark, $\Delta = 0$. If $\frac{d\Delta^m}{dp} > 0$, an increase in p^w will cause the $\Delta^m(\tau, p^w)$ locus to shift to the right, as shown in the right-side panel. Conversely, if $\frac{d\Delta^m}{dp} \leq 0$, an increase in p^w will cause the $\Delta^m(\tau, p^w)$ locus to shift to the *left*, as shown. It is immediately clear from the figures that the steady state outcome will be intimately linked to the sign of $\frac{d\Delta^m}{dp}$. Less obviously, the nature of transition dynamics will also hinge on the same condition, as we explore below.

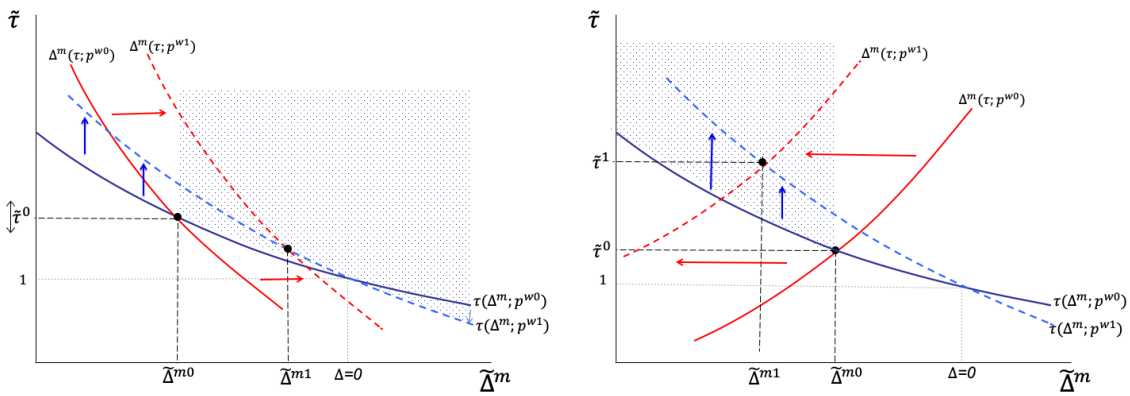


Figure 5: Steady State Response to $p^w \uparrow$ in (Δ^m, τ). [LHS: $\frac{d\Delta^m}{dp} > 0$, RHS: $\frac{d\Delta^m}{dp} \leq 0$]

4.1.2 Transition

We now describe the transition path from the original steady state to the new steady state following an unanticipated permanent increase in the terms of trade.

At the time of the shock, the distribution of human capital among the current voting population is fixed and given by voters' educational choices during youth under the original steady state at $t = T - 1$. That is, $\Delta_T^m = \tilde{\Delta}^{m0} = \Delta^m(\tilde{\mathbf{e}}^0)$. This serves as the relevant state variable that pins down the subsequent equilibrium sequence of tariff and education decisions, according to Definition 3.

While the young can adjust their educational decisions after the shock, the old who vote on trade policy cannot. Since the optimal tariff function at any given time depends on the *concurrent* value of Δ_t^m, $\tau_T = \tau(\Delta_T^m, p^{w1}) = \tau(\tilde{\Delta}^{m0}, p^{w1})$. Given our initial assumption that the returns to human capital are skewed toward the top ($\tilde{\Delta}^{m0} < 0$), the polarization result in Lemma 4 immediately implies that the equilibrium tariff will jump at the time of the shock, but will less than fully offset the increase in p^w:

Proposition 2. *Protectionist Surge: Starting from an initial steady state summarized by* $\{\tilde{\Delta}^{m0}, \tilde{\tau}^0; p^{w0}\}$ *where* $\tilde{\Delta}^{m0} < 0$ *and* $\tilde{\tau}^0 < \tau^P$*, an unanticipated increase in* p^w *at time* $t = T$ *will cause a concurrent increase in both the tariff and the domestic price relative to the initial steady state; i.e.* $\tau_T > \tilde{\tau}^0$ *and* $p_T > \tilde{p}^0$*, where* $\tilde{p}^0 \equiv \frac{p^{w0}}{\tilde{\tau}^0}$*. [Proof in Appendix B.2]*

Given the increase in the domestic relative price of the skill-intensive good at time T, we know from Lemma 1 that this increase would lead, ceteris paribus, to an increase in the educational investment of the young cohort born at time T relative to their predecessors. But at the same time, the young generation's educational decisions also depend on the expected price in the following period, and thus τ_{t+1}. Thus, the out of steady state education decisions for every member of generation T are given by $e_T(a) = e(a; \tau_T, \tau_{T+1})$ where the first argument, τ_T is already pinned down by $\tilde{\Delta}^{m0}$ but the second is endogenous and given by: $\tau_{T+1} = \tau(\Delta^m_{T+1}; p^{w1})$, which depends on generation T's educational decisions.

Under rational expectations, the equilibrium expected future tariff must coincide with the realized future tariff, which is a result of the political process in each subsequent period. The educational decisions of the young will shape future tariffs, while future tariffs determine young education decisions.[32] Our regularity assumption in (24) assures a unique fixed point solution to this problem in each period, so that transition is pinned down by parameters.

As intimated by Figure 5, there are two possibilities for how this transition will evolve depending on the underlying functional form assumptions. If young voters expect tariff liberalization (and therefore a higher skill premium), they will unambiguously acquire more education. But this expectation of liberalization will be realized only if these higher education levels allow the median voter to "catch up" to the overall economy enough that she will in fact be less protectionist in the future. This need not be the case. If despite an optimal educational response to the increase in the domestic skill premium at time T, the then-young median voter in generation T falls even further behind the overall economy so that $\Delta^m_{T+1} < \Delta^m_T$, then $\tau_{T+1} > \tau_T$: the median voter will even more protectionist following the shock.

We call the first possibility *Protectionist Overshooting*: following an initial tariff surge at the time of the shock, trade policy will gradually by liberalized as workers acquire more

[32] Note that under rational expectations, all agents must hold the same equilibrium beliefs about the future tariff. Given the assortative matching of initial advantage to optimal education levels and zero-mass voters, all agents understand that the median individual, a^m, will necessarily be the median voter with respect to trade policy in the subsequent period.

education, the distribution of human capital becomes less skewed, and protectionist pressures dissipate. Alternatively, in the case of *Protectionist Escalation* the initial tariff surge will be followed by a subsequent rise in tariffs, as workers become more politically polarized. (These two possibilities are separated by a knife-edge case, in which the tariff will jump immediately to the new steady state at the time of the shock)

We now show that whether the protectionist surge will be followed by gradual liberalization or a tariff escalation depends precisely on whether an increase in the skill premium causes convergence or divergence in the endogenous distribution of human capital; i.e. whether $\frac{d\Delta^m(p)}{dp} \gtrless 0$. In both cases, the terms-of-trade improvement triggers an immediate increase in the domestic skill premium and (thus) education levels at the time of the shock. But whether this increase in education exacerbates or mitigates polarization depends on the sign of $\frac{d\Delta^m(p)}{dp}$ and thus, ultimately, the concavity of the human capital function for different levels of a.

4.1.3 Protectionist Overshooting

Consider first the convergence case, which gives rise to *protectionist overshooting*. In this optimistic scenario, the increase in the skill premium will enable the median voter's human capital level to catch up with the rest of the population. As the median catches up her self-interested motive to raise tariffs is abated, and she will become less protectionist. The subsequently lower tariff will trigger future skill upgrading and catch up, leading to a reinforcing cycle of trade liberalization and skill upgrading. Thus, following the (inevitable) initial surge in protectionism at the time of the shock, the tariff will decline monotonically to a new steady state level that may (but need not) be below the initial steady state tariff. Formally:

Proposition 3. *Protectionist Overshooting. If* $\frac{d\Delta^m(p)}{dp} > 0$*, an unanticipated, permanent increase in* p^w *at time* T *leads to:*

i) an immediate increase in the tariff at time T*, from* $\tilde{\tau}^0$ *to* τ_T*;*

ii) a new steady state characterized by $\tilde{\tau}^1 < \tau_T$ *and* $\tilde{\Delta}^{m1} > \tilde{\Delta}^{m0}$*; and*

iii) a monotonic transition path after time T*,* $(\Delta^m_{T+t}, \tau_{T+t})$ $\forall t \geq 1$*, in which education increases and the tariff and polarization* (Δ^m) *decline and each period, converging to the new steady state.*

[Proof in Appendix B.3]

Figure 6 illustrates. The first panel shows the new steady state and transition in (e^m, τ) space. At the time of the shock, the tariff schedule pivots around the initial value of $\hat{e}(\tilde{\mathbf{e}}^{\mathbf{0}})$ to the locus $\tau^T(e^m)$. Since e^m is fixed for the current (old) median voter at $\tilde{e}^{m0}$, there is an immediate and unambiguous increase in the tariff to τ_T. Over time, the out-of-steady state education and tariff loci (not shown) will both gradually shift rightward as the domestic price and overall (mean) human capital level rise, leading to a monotonic decline in tariffs and increase in e^m to the new steady state.

The second panel offers the alternative depiction in (Δ_t^m, τ_t) space, which hews closely to the formal proof in Appendix B.3. Notice that the $\tau(\Delta_t^m; p^w)$ locus is the same in and out of steady state for any given p^w, since all of the arguments of the tariff function are contemporary to time t.[33] Conversely, because the time t older median voter's net skill position is a function of previous education decisions, the preceding period's tariff acts as a shift variable in the out-of-steady-state function $\Delta^m = \Delta^m(\tau_{t-1}, \tau_t; p^w)$. Lemma 3 implies that out-of-steady-state $\Delta^m(\tau_{t-1}, \tau_t; p^w)$ schedules are steeper than the steady state schedule, $\Delta^m(\tau; p^w)$ for any given p^w.[34] In the overshooting case, the $\Delta^m(\cdot)$ schedules are downward sloping as shown. For any given p^w, the steady state equilibrium is given by the intersection of the steady state loci, while Assumption 2 ensures that the steady state schedule $\Delta^m(\tau; p^w)$ locus intersects $\tau(\Delta^m; p^w)$ only once and from above.

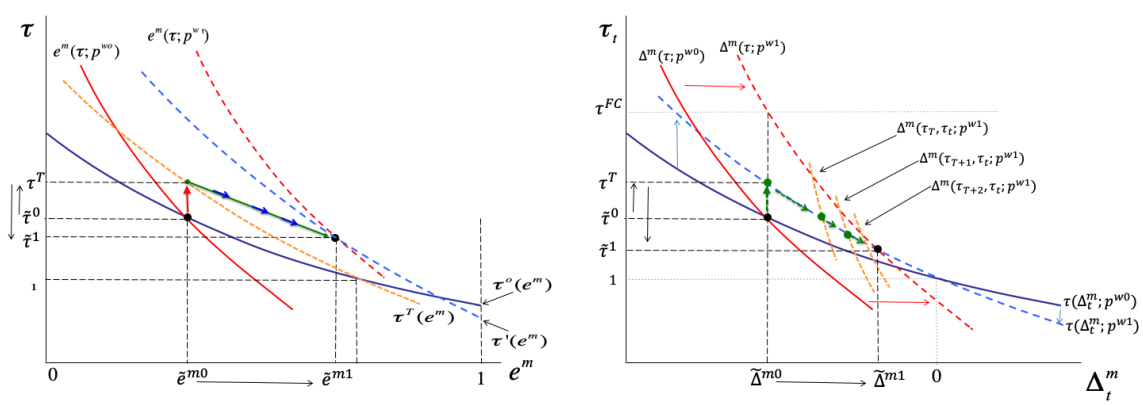


Figure 6: Protectionist Overshooting

[33] $\tau(\Delta_t^m; p^w) = \tau(\Delta^m; p^w)$ iff $\Delta_t^m = \Delta^m$.

[34] Intuitively, Δ^m is less responsive to changes in either the contemporary or past tariff than it is to a change in both tariffs together; Claim 7 in Appendix C formalizes.

Following the terms-of-trade shock at time T, the equilibrium time path is defined as the series (Δ_t^m, τ_t) where for each period $t > T$, $\Delta_t^m = \Delta^m(\tau_{t-1}, \tau_t; p^{w1})$ intersects the (new) steady state tariff locus, $\tau(\Delta_t^m; p^{w1})$. Thus, starting from the initial steady state $(\tilde{\Delta}^{m0}, \tilde{\tau}^0)$, an unanticipated increase in the terms of trade from p^{w0} to p^{w1} at time T causes the tariff locus to pivot clockwise around $\Delta^m = 0$ from $\tau(\Delta_t^m, p^{w0})$ to $\tau(\Delta_t^m, p^{w1})$, according to Lemma 4. (That is, every initially protectionist voter will become more protectionist at the time of the terms-of-trade shock.) The steady state $\Delta^m(\tau; p^w)$ locus shifts to the right from $\Delta(\tau, p^{w0})$ to $\Delta(\tau, p^{w1})$, since for any given tariff level, higher world prices will eventually induce educational investment and income convergence (in the overshooting case).

At the time of the shock, the old generation cannot adjust their educational choices, (i.e. $\Delta_T^m = \tilde{\Delta}^{m0}$), and so the tariff jumps immediately to $\tau_T > \tilde{\tau}^0$ as shown (Proposition 2). Equilibrium at time $T+1$ is then given by the intersection of the new tariff locus, $\tau(\Delta_t^m; p^{w1})$, and $\Delta^m(\tau_T, \tau_{T+1}; p^{w1})$. As we show in the proof, the out-of-steady-state $\Delta^m(\tau_T, \tau_{T+1}; p^{w1})$ function coincides with the (new) steady state $\Delta^m(\tau; p^{w1})$ at τ_T, as shown; thus, it must hold that $\tau_{T+1} < \tau_T$ (and $\Delta_{T+1}^m > \Delta_T^m$). Since $\tau_{T+1} < \tau_T$, the next period's schedule, $\Delta^m(\tau_{T+1}, \tau_{T+2}; p^{w1})$, must lie strictly to the right of the previous schedule, resulting in a yet-lower tariff at $T + 2$t. Each period thereafter, the out of steady state $\Delta^m(\cdot)$ schedule continues to shift right, gradually converging to the new steady state along the new steady state tariff locus.

Figure 7 maps the time path of the equilibrium tariff in this overshooting case. The new steady state tariff level may be higher or lower than the original steady state; absent additional assumptions it could go either way. Regardless, the policy overshooting result obtains: there is an immediate surge in protectionism following an exogenous terms-of-trade shock, followed by a gradual decline in tariffs as the new steady state tariff level is reached. Even if a terms-of-trade shock will ultimately result in lower tariffs, the short run response points in exactly the opposite direction: even a "rosy" long run is preceded by a rocky transition.

Crucially, the non-monotonicity depicted in Figure 7 hinges on both inequality in the returns to trade and sticky labor market adjustment. If instead voters were identical, they would have no self-interested motive to distort prices, and thus would always choose free trade, regardless of the world price.[35] Or alternatively, if economic adjustment were

[35] In this small open economy, free trade maximizes national income, and therefore the indirect utility for every economically representative ("average") voter. In a large country, the national income maximizing tariff would be given instead by the inverse elasticity of foreign export supply, per Johnson (1952).

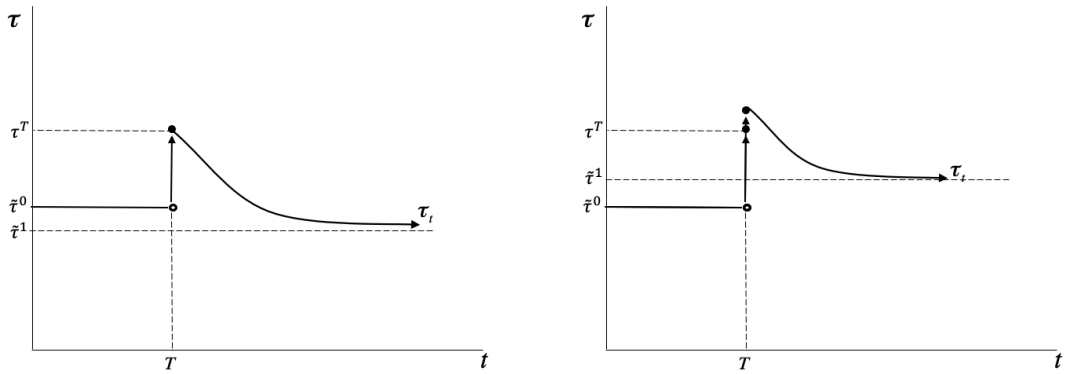


Figure 7: Overshooting Tariff Response to $p^w \uparrow$ if $\frac{d\Delta^m}{dp} > 0$.

immediate, the economy would simply jump to the new steady state at the moment of the terms-of-trade shock. It is therefore specifically the *combination* of inequality and labor market stickiness that generates the rich political-economy transition dynamics presented here.

Finally, viewing the transition dynamics through the lens of domestic prices reveals that the protectionist surge at time T is acting as a shock absorber for the overall economy. As we see in Figure 8, the sudden, sharp political response to the increase in world prices tempers the immediate effect of the shock on local prices, which effectively gives the country's constituents time to adjust gradually to the new macroeconomic conditions. This gradual adjustment in education level is depicted by the right hand side panel of Figure 8.

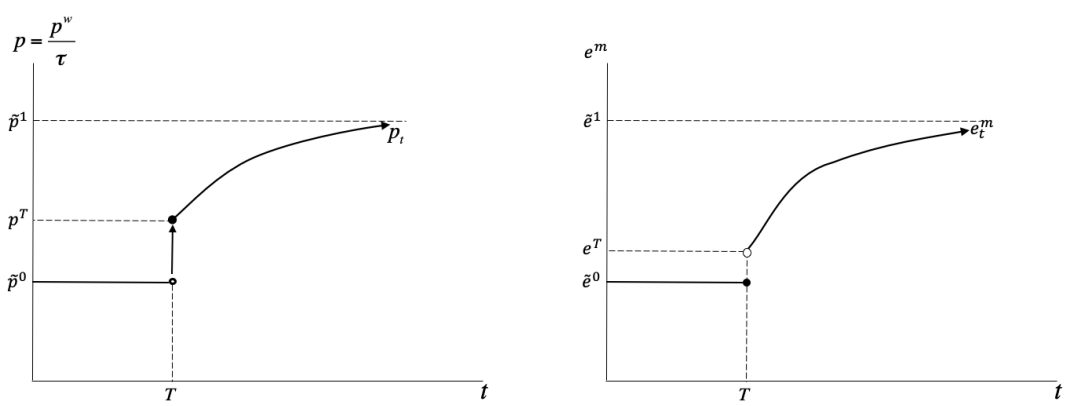


Figure 8: Overshooting Time Path for Prices and Education

Protectionist overshooting is not innocuous. The surge in the tariff at time T slows subsequent human capital acquisition for generations and thus entails real efficiency losses. From a utilitarian perspective, the economy would be better off if it could immediately shift to the new steady state at time T. Section 5 explores the potential for welfare-improving policy interventions to mitigate the initial tariff surge or speed the pace of adjustment.

4.1.4 Protectionist Escalation

We now turn to examine the alternative case, in which a rising skill-premium exacerbates the underlying inequality. In this case the initial surge is followed by further increases in the level of protection. Formally, we have:

Proposition 4. *Protectionist Escalation. If* $\frac{d\Delta^m(p)}{dp} \leq 0$*, an unanticipated, permanent increase in* p^w *at time* T *leads to:*

i) an immediate increase in the tariff at time T*, from* $\tilde{\tau}^0$ *to* τ_T*;*

ii) a new steady state characterized by $\tilde{\tau}^1 \geq \tau_T > \tilde{\tau}^0$ *and* $\tilde{\Delta}^{m1} < \tilde{\Delta}^{m0}$*; and*

iii) a transition path $(\Delta^m_{T+t}, \tau_{T+t}) \forall t \geq 1$*, that oscillates around and converges to the new steady state. (In the razor's edge case in which* $\frac{d\Delta^m(p)}{dp} = 0$*, transition will be instantaneous at time* T*.)*

[Proof in Appendix B.4]

Figure 9 illustrates. Again, the first panel shows the new steady state and transition in (e^m, τ) space. Following the initial terms-of-trade shock, the time T steady state locus pivots clockwise to $\tau^T(e^m)$, in orange. After time T, the out-of steady state education and tariff loci (not shown), shift toward the new steady locus, following an oscillating convergence pattern. Over time, both loci shift toward the new steady state as the tariff swings gradually dissipate.

The second panel depicts the same mechanics in (Δ_t, τ_t) space, again in close parallel with the formal proof. Notice that in this escalation scenario, polarization is exacerbated by a higher domestic relative price, p (reduced by a higher tariff): i.e. $\frac{d\Delta^m}{dp} \leq 0$ ($\frac{d\Delta^m}{d\tau} \geq 0$).[36] For any tariff level, the increase in p^w will eventually amplify the polarization of wages as

[36] Lemma 3 again implies that the out-of-steady-state $\Delta^m(\tau_{t-1}, \tau_t; p^w)$ schedules are steeper than the steady state $\Delta(\tau; p^w)$ schedule, as shown.

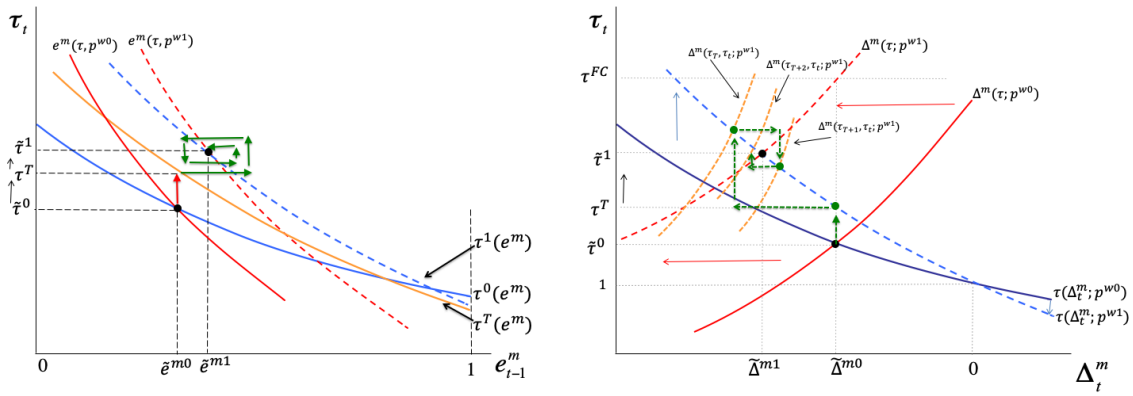


Figure 9: Protectionist Escalation

rising education levels let the median voter fall behind the mean. Thus the new steady state $\Delta(\cdot)$ schedule lies strictly to the left of the initial schedule. As in the overshooting case, the terms-of-trade shock causes the tariff locus to pivot clockwise around $\Delta = 0$, further polarizing tariff preferences immediately.

At the time of the shock, however, education is fixed, so that $\Delta^m_T = \tilde{\Delta}^{m0}$, and so the tariff jumps to τ_T. At $T+1$, the new equilibrium is given by the intersection of the new tariff schedule and $\Delta^m(\tau_T, \tau_t; p^{w1})$. Since the tariff is less than fully compensating, the domestic prices rises (despite the increase in the tariff) and $p_T > \tilde{p}^0$, which will cause voters to increase their education levels, causing the $T+1$ locus $\Delta^m(\tau_T, \tau_t; p^{w1})$ to shift left, as shown. In this scenario, the increase in edcuation exacerbates polarization, so that $\Delta^m_{T+1} < \Delta^m_T$ and therefore $\tau_{T+1} > \tau_T$. The next period will see a swing in the opposite direction. Because the time $T+1$ tariff is higher than it was at T, the domestic price will fall somewhat ($p_{T+1} < p_T$), which will dampen polarization , and therefore and protectionism. This shift will allow the $\Delta^m(\tau_{T+1}, \tau_t; p^{w1})$ to shift back toward the right as shown. Convergence proceeds by oscillation: when the tariff rises, inequality falls, which pushes the subsequent tariff lower; the lower tariff then causes inequality to rise again (though not so much as to offset the previous decline), which causes the next period's tariff to rise, but not all the way to its previous level.

The case of protectionist escalation highlights an uncomfortable political tension that arises when education and inequality move together. When a boost in the domestic skill premium induces educational investments that *increase* economic polarization (i.e. $\frac{d\Delta^m}{dp} < 0$), a terms-of-trade improvement will exacerbate inequality and the tariff will continue to

rise even after the initial surge in trade protection. Although the political pendulum will alternately swing the other way, and eventually the swings in tariffs and human capital will moderate as they converge to the new steady state, the transition will be politically tempestuous. In a multilateral world, such tariff swings could be highly disruptive for a rules-based trading system or other institutional norms based historical reciprocity.

4.2 Technology Shocks

It is straightforward to show that the political implications of skill-biased technological change (SBTC) can mimic the effects of a terms-of-trade shock. Consider the effect of a permanent, unanticipated increase in the relative productivity of skilled labor, summarized by the parameter b in our model. The following proposition establishes that, although the underlying mechanics are different, the political effects of a technological shock are qualitatively similar to the effects of the terms-of-trade shock explored above.

Proposition 5. *Polarizing effect of SBTC. Starting from an initial steady state summarized by* $\{\tilde{\Delta}^{m0}, \tilde{\tau}^0; p^{w0}\}$ *where* $\tilde{\Delta}^{m0} < 0$ *and* $\tilde{\tau}^0 < \tau^P$*, an unanticipated skill-augmenting technological improvement that increases* b^0 *to* $b^1 > b^0$ *at time* $t = T$ *leads to:*

i) an immediate increase in the tariff at time T*, from* $\tilde{\tau}^0$ *to* τ_T*,*

ii) followed subsequently by either:

(a) if $\frac{d\Delta^m}{dp} \geq 0$*, a monotonic decline in the tariff to a new steady state* $\tilde{\tau}^1 \leq \tau_T$*; or*

(b) if $\frac{d\Delta^m}{dp} < 0$*, oscillating convergence to a more protectionist steady state,* $\tilde{\tau}^1 > \tau_T$*.*

[Proof in Appendix B.5]

Intuitively, an increase in b at time T immediately magnifies the then-old median voter's initial self-interested motive to manipulate the domestic price, since $\Delta^m_T = \Delta^m(a^m, \tilde{\mathbf{e}}^{\mathbf{0}}; b^1)$ $= (1-\alpha)b^1[h(a^m, \tilde{e}^0(a^m)) - \bar{h}(\tilde{\mathbf{e}}^{\mathbf{0}})] > \tilde{\Delta}^0$. Over time, education will respond to both the exogenous technology shock and the endogenous evolution of tariffs, following the earlier logic.

While the speed and magnitude of terms-of-trade shocks and SBTC surely differ in practice – plausibly, the 'China shock' may have been faster than the advance of labor-replacing technology – the political implications may be commensurate. By increasing the

relative demand for import-competing labor, tariffs offer adversely-affected workers a policy tool with which to tilt labor demand in their favor.

There is a heated debate about whether technological change or import competition (especially from China) bears greater responsibility for the recent labor market polarization in the US and elsewhere;[37] Proposition 5 suggests that the root cause may be politically immaterial. Whether caused by technology or trade, rising economic inequality may have the same political consequences for trade policy. Globalization may simply be technology's scapegoat.

5 Discussion

This section uses the model to discuss how domestic policies and multilateral trade rules may defuse or exacerbate protectionist pressure in the long and short run.[38] We begin by asking whether introducing (exogenous) domestic redistribution or education policies to our political economy model could mitigate voters' use of tariffs. Turning to multilateral policy, we then revisit the case for escape clause (or 'safeguard' provisions) in trade agreements in the context of our model. Based on the theory, we make five main points.

1. Popular support for protectionism falls when individual voters' incentives are more closely aligned with the overall economy. Recall that the first order condition in (15) implicitly defines the most preferred tariff for a voter as a function of her relative net skill position:

$$(\tau_t - 1)\tau_t = \frac{\Delta(a)}{d\bar{E}_t/d\tau_t} \geq 0 \iff \Delta(a) \leq 0 \tag{26}$$

From this expression, is it clear that any policy that seeks to reduce popular pressure to implement a tariff must reduce or offset the magnitude of individual self-interest $|\Delta(a)|$ for a sufficiently large set of politically decisive voters.[39] As long as some part of individuals' earnings are derived from market wages, and as long as there is underlying inequality in

[37] See, e.g., Goos and Manning (2007), Autor and Dorn (2013), Goos, Manning, and Salomons (2014), Autor, Dorn, and Hanson (2015)

[38] In this exercise, we are effectively stepping outside of a strict median voter framework to adopt the perspective of a social planner who is designing domestic economic institutions or multilateral trade rules subject to the condition that voters will choose trade policy endogenously. Extending political economy models to include multiple endogenous policy tools remains a challenge unless one is willing to collapse the policy set to a single dimension.

[39] One need not "buy off" all voters in a democracy, but just enough to swing the election.

the distribution of those market wages, voters will have an incentive to sacrifice at least a little bit of aggregate income in order to tilt the wage distribution in their favor.

It follows that any *unconditional* (net of tax) redistribution program where payments are divorced from wages – including a universal basic income scheme – may have a limited influence on trade policy preferences. Indeed, if transfer payments are completely independent of prices (even if highly progressive in a), a literal interpretation of our model would imply that the transfer scheme would have *no* effect on the optimal tariff, since the transfers would not even enter the optimal tariff expression in (26).[40] Alternatively, if transfers depended on prices but not individual characteristics (for instance, via aggregate national income), they would add another term to the right-side denominator in (26) and therefore reduce – but not eliminate – the influence of self-interest in tariff preferences.[41] More generally, any tax and transfer scheme would need to depend on both domestic prices and a to exactly offset $\Delta(a)$ in order to completely eliminate individual self-interest from influencing tariffs. This seems unlikely.

2. Conditional redistribution policies that successfully reduce the role of individual self-interest in tariff preferences generally would also blunt young voters' incentives to acquire education. For instance, a progressive tax and transfer scheme tied to market wages would reduce the dispersion in post-tax earnings, and therefore protectionist pressure, but it would also reduce human capital acquisition, especially at the top, and thus aggregate income.[42] Our model thus highlights a fundamental tension between economic efficiency and politics: to defuse protectionist pressure, a policy intervention needs to enter the first order condition governing individuals' tariff preferences in (15), but not the first order condition governing their optimal educational decisions in (8). That is a tall order.

Educational subsidies come close to this ideal, since they can both increase individuals' incentive to acquire education, and, if targeted to increase $\Delta(a^m)$ closer to zero, simultaneously defuse populist pressure to raise tariffs. But they are not costless. Financing subsidies to education requires tax revenue. If collected lump sum, the tax would be regressive. If financed instead through progressive taxation, the effect would be to distort downward

[40] The exception is the limiting case in which the income payment entirely replaces individual income.

[41] To the extent that unconditional redistribution increases workers' labor market flexibility, such a scheme may reduce the dispersion of $\Delta(a)$, which would reduce both inequality and protectionist pressure. Recent research by Bryan, Chowdhury, and Mobarak (2014) suggests that even small cash transfers can be a powerful tool for increasing workforce flexibility and alleviating poverty. Our theory suggests that these same forces therefore could have important dynamic political consequences as well.

[42] Willmann (2004) shows that such an investment disincentive can overturn the gains from trade.

educational attainment at the top, reducing both economic efficiency and the country's comparative advantage. In practice, many public investments in education already accrue to the top, particularly in the US (for instance, tax credits for higher education). Absent fundamental structural reforms, simply increasing public spending on education therefore could exacerbate underlying inequality, and thus protectionist pressure, consistent with our previous findings in a static setting in Blanchard and Willmann (2016).

3. What matters in the long run is whether or not less-skilled workers are able to catch up to the overall economy. The long-run consequence of the macroeconomic shock, whether the tariff eventually will fall via overshooting or rise via escalation, hinges on the sign of $\frac{d\Delta(p)}{dp}$. If a rising skill premium induces *and enables* individuals at the bottom of the income distribution to increase education enough that their human capital levels start to catch up with the rest of the economy, so that $\frac{d\Delta(p)}{dp} > 0$, then as soon as workers have an opportunity to move up the educational ladder, the initial protectionist surge will begin to reverse and inequality will decline. But if instead $\frac{d\Delta(p)}{dp}(p) < 0$, then inequality, and thus the demand for tariff protection, will continue to rise after the initial protectionist surge – even though education will be uniformly higher than in the initial steady state.

Domestic economic policies can influence the sign of $\frac{d\Delta(p)}{dp}$. Starting from a protectionist surge, progressive reforms in education would speed convergence to a new, lower steady state tariff. Conversely, spending cuts that reduce opportunities for low-skilled workers or regressive changes in the tax code would have the opposite effect, and could even shift the long run equilibrium from protectionist overshooting to escalation.

4. Reducing labor market frictions among voters at the time of the shock will speed transition to the new steady state. The obvious but important implication is that increasing voter turnout among the younger generations would speed transition. Alternatively, if older agents had access and the incentive to acquire education in the second stage of life, the initial protectionist surge would be smaller and transition to the new steady state would be faster, because the old would act "younger" in the model, increasing their skills in response to the initial terms-of-trade shock (or SBTC). More generally, reductions of the many frictions that limit workers' ability to respond to a changing national labor market has the potential to increase support for globalization.

5. Turning from domestic policy implications to multilateral, our model offers an argument in support of including escape clauses (safeguards) in trade agreements. Consider the case of protectionist overshooting depicted in Figure 6, in which an initial protectionist spike is a

temporary response to an unanticipated shock, and the new steady state tariff is below the initial steady state $\tilde{\tau}^1 < \tilde{\tau}^0$. In this scenario, the 'home' country's tariff would eventually fall below the original steady state level if allowed to run its course, leaving both this country and (if it is large in world markets) its trading partners better off.[43]

Absent a safeguard provision, however, this adjustment path may not have room to play out. Without an escape clause that allows countries the opportunity to temporarily raise tariffs in response to shocks, the initial protectionist surge likely would be met by tariff retaliation. While in some circumstances the threat of retaliation could deter short run tariff surges, there is nothing to stop an adversely-affected median voter from starting a trade war.[44] An increase in foreign tariffs would in turn worsen the home country's terms of trade (lowering p^w) resulting in higher 'new' steady state tariff (both the education and tariff loci would shift back toward the initial steady state). Thus, in the absence of safeguards, a short-term protectionist spike could lead to permanent protection: the opportunity to reach $\tilde{\tau}^1$ would be lost.

6 Empirical Context

Our theoretical analysis highlights several insights that can be used to assess the potential for protectionist surges and their long-run consequences. First, greater dispersion in individuals' returns to economic openness will result in larger and longer-lasting protectionist reactions to a terms-of-trade improvement or skill-biased technological change. Second, economic mobility – how long it takes for workers to adjust to macroeconomic shocks – is a crucial determinant of longer term welfare. And third, conditional on a protectionist surge, the model suggests that whether the long run outcome will be protectionist overshooting or further tariff escalation depends on whether economic polarization is increasing or decreasing as overall educational achievement increases.

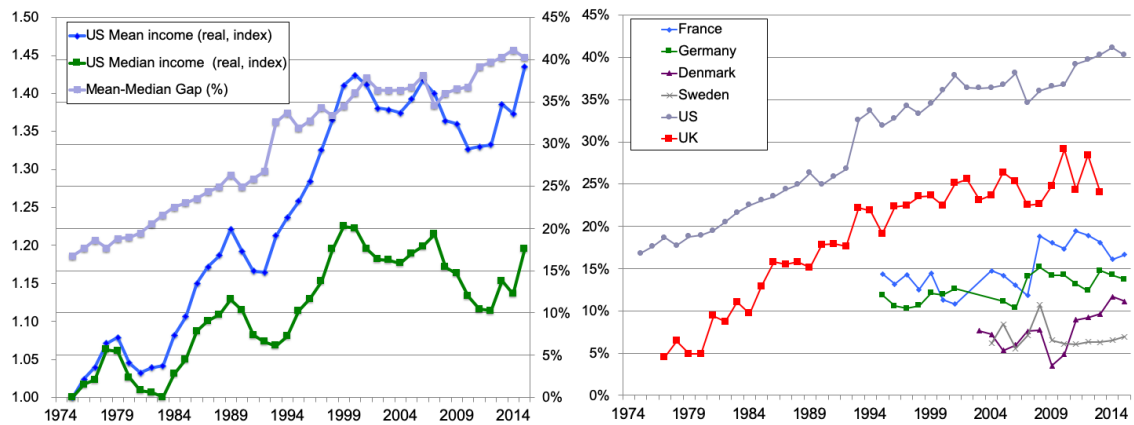
Below, we review recent evidence on these three potential drivers of protectionism indicated by the model. Although each of our measures is unquestioningly imperfect, to-

[43] Relaxing the small country assumption, a decline in the 'home' country tariff would imply a terms-of-trade improvement for its trading partners.

[44] Indeed, one can prove that if the potential trade war could be guaranteed to be sufficiently small, the median voter would choose to trigger the fight: she will not incur the long run consequences, and in the short run, she stands to gain from a marginal decline in p^w. A sufficiently large trade war, however, could leave her worse off, and thus could be an effective deterrent.

gether, they offer an opportunity to evaluate the practical potential for protectionist surges through the lens of theory and disciplined by contemporary data.

Rising Inequality. The model predicts that protectionism is increasing with the gap between the median voter's income and the average.[45] Here, we proxy for Δ^m using the percentage difference between mean and median (pre-tax) household income, and compare these mean-median gaps across countries and over time. Figure 10 uses data from the US Census, the UK Office of National Statistics, and EuroStat to compare the gap between mean and median household (pre-tax) income over time and across countries.



Source: US Census; UK ONS; Eurostat

Figure 10: Income Gap Across Countries and Over Time

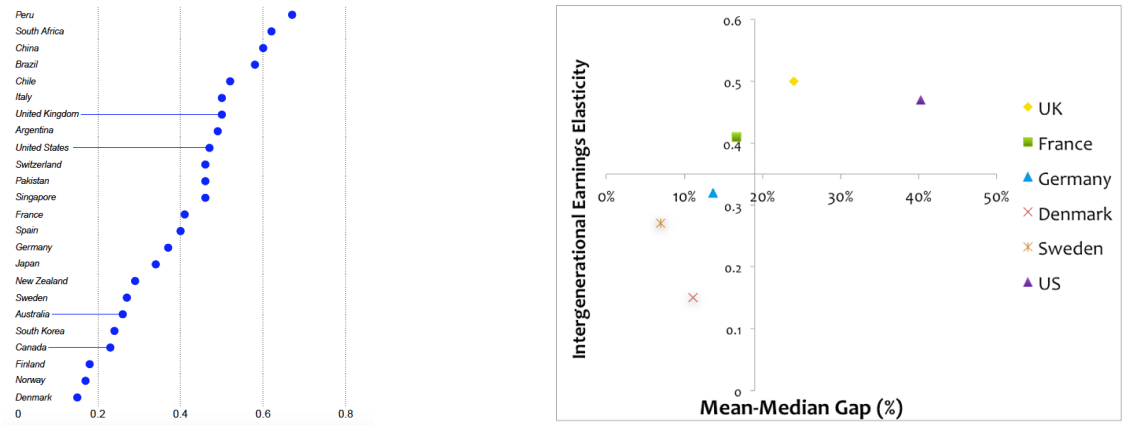
The leftmost panel charts the change in US real household mean and median incomes, indexed to 1974. The mean-median gap (defined as the difference between mean and median income as a share of median income) has risen steadily since the start of the period, roughly doubling over the course of 40 years. The right-side panel of Figure 10 then compares the mean-median gap across countries and over time.[46] It is clear from the figure that by this measure, inequality in the US and the UK has been rising systematically over the past four decades and is demonstrably higher than in other wealthy countries.

Labor Market Frictions. Measuring labor market frictions is notoriously difficult, particularly across countries and over time. Intergenerational income elasticity (IGE) estimates from Corak (2006) offer a rough proxy, reflecting differences in long run economic

[45] In the model, the difference between the median voter's income and average income at any time t is given by: $I_t^o(a^m; e_{t-1}(a^m)) - \int_a I_t^o(a, e_{t-1}(a))f(a)da = b[h(a^m, e_{t-1}(a^m)) - \bar{h}(a, \mathbf{e_{t-1}}(a))] = \frac{1}{1-\alpha}\Delta_t^m$.

[46] Data are shown for all years available from these sources.

mobility across countries. The left-side panel of Figure 11 reproduces the cross-sectional IGE measures in Corak (2006), where higher values indicate greater persistence in income levels across generations, which suggests lower income mobility.[47] The right-side panel then combines these IGE estimates with the most recent measures of the mean-median income gap for each country in Figure 10, where we normalize the axes at the mean value of each indicator in the sample shown.[48] The upshot is immediate: by these (admittedly rough) proxies, the US and UK are outliers: less mobile and less equal than otherwise comparable developed countries.



Source: Corak (2006); US Census; UK ONS; Eurostat

Figure 11: Inequality and Economic Mobility across Countries

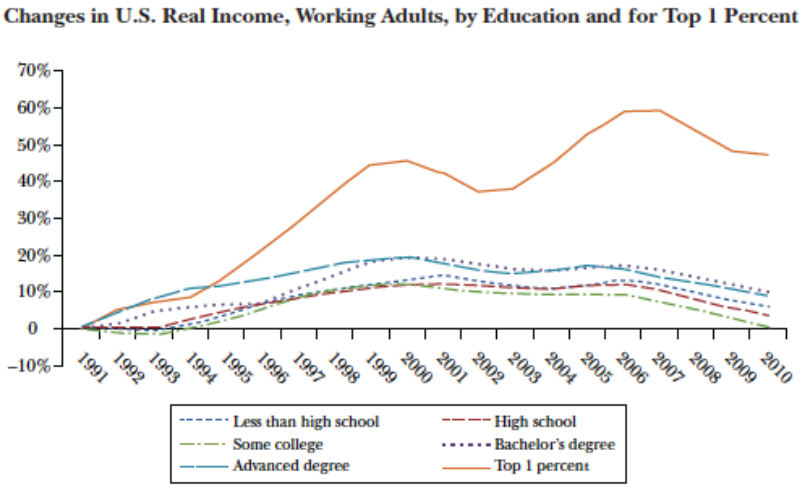
Catching up or Falling Behind? The starkest prediction of our model is the potential for two very-different long-run outcomes following a protectionist surge, depending on whether inequality rises or falls over the long run. Importantly, the model also predicts an unequivocal increase in education over the long term in response to terms-of-trade changes that depress wages for less-skilled workers and/or skill-biased technological advances that increase the returns to human capital directly. Indeed, by virtually any measure there has been a near-universal increase in educational achievement in developed and developing countries over the past half-century.[49]

[47] IGE measures *realized* income mobility, whereas ideally we would want to measure *potential* income mobility. We are not aware of reliable cross-country measures of the latter, however.

[48] This same pattern can be seen in the now-famous "Great Gatsby Curve", introduced in a 2012 speech by then-chair of the Council of Economic Advisors, Alan Krueger. Both depictions use the IGE estimates from Corak (2006) to measure social mobility, but Krueger used GINI coefficients to measure inequality rather than the mean-median gap that is directly indicated by our model.

[49] See, e.g. Barro and Lee (2013).

But rising educational achievement does not mean that income or the distribution of the returns to human capital is becoming less skewed.[50] Goldin and Katz (2009) identify the simultaneous rise in both educational achievement and the skill premium in the US in recent decades, while Castelló-Climent and Doménech (2014) demonstrate similar patterns for a broad set of 146 countries since 1950. Despite a secular increase in education, these studies demonstrate that the skill premium has continued to rise, particularly for the very top ("superstar") income earners. Indeed, Haskel, Lawrence, and Slaughter (2012) find that US workers with the median level of education (which falls in the category of "Some College") experienced both the lowest real income growth from 1991 to 2012, and the steepest decline since 2000. We replicate their figure with permission in Figure 12. Taken together, these data suggest that income inequality may continue to grow *despite* rising levels of educational attainment. The implication in the context of our model is sobering: if indeed $\frac{d\Delta^m}{dp} < 0$ as these data seem to suggest, protectionism may continue to escalate.



Source: US Census via Haskel, Lawrence, and Slaughter (2012)

Figure 12: Education does not guarantee rising income

In summary, the data on inequality, economic mobility, and the increasingly polarized returns to human capital despite educational advances paint a grim picture. Our theoretical analysis suggests that these economic factors may play a role in recent anti-globalization political shifts. On the flip side, it is worth noting that widespread popular support for global economic engagement during the middle of the twentieth century coincided with a dramatic

[50] Moreover, even a reduction in the skewness in the distribution of education is not enough to ensure a reduction in the skewness of the *market returns* to human capital.

increase in middle class incomes and economic mobility following the end of the second world war. But do not take this too far. Economics is one of many drivers of electoral outcomes, and trade policy rarely plays a central role in major elections. Nonetheless, our theory offers a new way to approach data when thinking about the long and short run consequences of recent changes in technology and trade. More broadly, the model highlights how rising economic polarization exacerbates individuals' incentives to use distortionary policy tools to tilt the distribution of market wages in their favor. If inequality continues to rise, theory suggests that populist support for market interventions like tariffs is likely to increase.

7 Concluding Remarks

We develop a tractable dynamic political equilibrium model to identify the role of inequality and labor market frictions in shaping democratic political responses to macroeconomic shocks over time. In our model, the extent of trade protection depends on the underlying distribution of human capital, and hence the distribution of the gains from trade. Unanticipated trade or technology shocks that exacerbate underlying inequality will lead to a short-run surge in protectionism: when policy can change faster than workers can adjust, trade policy serves as a country's 'shock absorber.'

We show that the long-run consequences of a shock depend on whether or not less-skilled workers are eventually able to catch up to the overall economy. If convergence is possible, the result will be *protectionist overshooting*: the short run tariff spike will gradually unwind, as workers increase education and support for freer trade rises. Alternatively, if less-skilled workers fall even further behind after the shock, the result will be *protectionist escalation*: a pendulous transition to permanently higher tariffs.

We use the model to construct a set of criteria for evaluating the likely political implications of education and redistribution policies in the short and long run. The exercise highlights a tension between economic efficiency and politics: optimal investment in human capital requires strong individual incentives, but ex-post inequality in the gains from trade can lead to political distortions that are costly in the long run. Finally, we present data on economic mobility and income inequality that suggest the US and UK are outliers relative to other OECD countries, with relatively low economic mobility and high inequality. Evidence of continued polarization in the returns to human capital suggest that protectionism may continue to escalate despite rising overall educational attainment.

Continuing in the context of trade and economic nationalism, our model can be extended to incorporate endogenous voter turnout in response to macroeconomic shocks. For instance, to the extent that macroeconomic shocks polarize the political preferences of the electorate, they may increase voter turn-out at the extremes, potentially leading to large (or volatile) policy swings. Along a different line, our model can be extended to evaluate the opportunity for intergenerational rent transfers: to what extent could the young 'buy off' the old in an effort to reduce or eliminate protectionist surges? Under what circumstances could a time-consistent constitutional agreement or transfer scheme prevent overshooting or to reverse escalation? Or, in a model with hereditary or autocorrelated advantages over generations, how does the ability to pass-on initial advantages to one's children affect short- and long-term political responses to changing technology or trade?

Finally, as we emphasized in the introduction, our approach can be used to explore a wide variety of policy questions beyond trade. Although we make specific assumptions to focus on the recent rise of the anti-globalization movement, the basic theoretical insights and mechanisms at the core of our theoretical analysis are germane to a wide set of political economy applications. In our model, economic adjustment takes place through human capital acquisition and politics are determined by majoritarian voting, but both can be understood as representing a broader class of possibilities. Economic adjustment could instead take the form of physical capital accumulation, changes in land use, technology adoption, or pension saving. Likewise, one could incorporate a host of alternative political decision rules in which, at least to some extent, distributions matter. Accordingly, the basic overshooting insight – that differential friction between economic and political change can drive policy overshooting that retards long run adjustment to shocks – is transportable to a host of alternative contexts, including adoption of new technologies, pension reforms, and political responses to climate change.

References

Acemoğlu, D., G. Gancia, and F. Zilibotti (2015). Offshoring and Directed Technical Change. *American Economic Journal: Macroeconomics* 7(3), 84–122.

Acemoğlu, D., S. Naidu, P. Restrepo, and J. Robinson (2015). Democracy, Redistribution, and Inequality. In A. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2B, Chapter 21, pp. 1885–1906. Elsevier. Working Paper for the Handbook of Income Distribution.

Acemoğlu, D. and J. A. Robinson (2013). Economics versus Politics: Pitfalls of Policy Advice. *Journal of Economic Perspectives*.

Alesina, A. and D. Rodrik (1994). Distributive Economics and Economic Growth. *Quarterly Journal of Economics 109*(2), 465–490.

Antras, P., A. deGortari, and O. Itskhoki (2016). Globalization, Inequality, and Welfare. Working Paper 22676, NBER.

Artuç, E., S. Chaudhuri, and J. McLaren (2010). Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review*.

Autor, D. and D. Dorn (2013). The Growth of Low Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review 103*(5), 1553–1597.

Autor, D., D. Dorn, and G. Hanson (2013). The China Syndrome: Local Labor Effects of Import Competition in the United States. *American Economic Review 103*(6), 2121 – 2168.

Autor, D., D. Dorn, and G. Hanson (2015). Untangling Technology and Trade: Evidence from Local Labor Markets. *The Economic Journal 125*(584), 621–646.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review*. forthcoming.

Barro, R. and J.-W. Lee (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics 104*, 184–198.

Bassetto, M. (1999). Political Economy of Taxation in an Overlapping-Generations Model. *Federal Reserve Bank of Minneapolis Discussion Paper no. 133*.

Blanchard, E. and G. Willmann (2011). Escaping a Protectionist Rut: Policy Mechanisms for Trade Reform in a Democracy. *Journal of International Economics 85*(1), 72–85.

Blanchard, E. and G. Willmann (2016). Trade, Education, and the Shrinking Middle Class. *Journal of International Economics 99*(1), 263–278.

Brainard, L. and T. Verdier (1997). The Political Economy of Declining Industries: Senescent Industry Collapse Revisited. *Journal of International Economics 42*, 221–38.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014, September). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica 82*(5), 1671–1748.

Castelló-Climent, A. and R. Doménech (2014, March). Human capital and income inequality: Some facts and some puzzles. Working Paper 12/28, BBVA.

Chetty, R., N. Hendren, and L. Katz (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review 106*(4), 855–902.

Corak, M. (2006). Do Poor Children Become Poor Adults? Lessons for Public Policy from a Cross Country Comparison of Generational Earnings Mobility. *Research on Economic Inequality 13*.

Dhingra, S. (2014). Reconciling Observed Tariffs and the Median Voter Model. *Economics and Politics 26*(3), 483–504.

Dix-Caneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica 82*(3), 825–885.

Dornbusch, R. (1976). Expectations and Exchange Rate Dynamics. *Journal of Political Economy 84*, 1161–76.

Dutt, P. and D. Mitra (2002). Endogenous Trade Policy Through Majority Voting: A Empirical Investigation. *Journal of International Economics 58*, 107–133. Roughly, an empirical test of Mayer 1984.

Fernandez, R. and D. Rodrik (1991). Resistance to Reform: Status Quo Bias in the Presence of Individual Specific Uncertainty. *American Economic Review 81*, 1146–1152.

Gawande, K. and U. Bandyopadhyay (2000). Is Protection for Sale? Evidence on the Grossman-Helpman Theory of Endogenous Protection. *Review of Economics and Statistics 82*(1), 139–152.

Gennaioli, N. and G. Tabellini (2019). Identity, Beliefs, and Political Conflict. CESifo Working Paper 7707.

Goldberg, P. K. and G. Maggi (1999). Protection for Sale: An Empirical Investigation. *American Economic Review 89*(5), 1135–1155.

Goldin, C. and L. Katz (2009). The Race Between Education and Technology: The Evolution of US Wage Differentials, 1890-2005. NBER Working Paper 12984.

Goos, M. and A. Manning (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *Review of Economics and Statistics 89*, 118–33.

Goos, M., A. Manning, and A. Salomons (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review forthcoming*.

Grossman, G. M. and E. Helpman (1994). Protection for sale. *American Economic Review 84*(4), 833–850.

Grossman, G. M. and E. Helpman (2020). Identity Politics and Trade Policy. *Review of Economic Studies*. forthcoming.

Haskel, J., R. Lawrence, and M. Slaughter (2012). Globalization and U.S. Wages: Modifying Classic Theory to Explain Recent Facts. *Journal of Economic Perspectives 26(2)*, 119–40.

Hassler, J., J. V. Rodríguez Mora, K. Storesletten, and F. Zilibotti (2003). The Survival of the Welfare State. *American Economic Review 93*, 87–112.

Hassler, J., K. Storesletten, and F. Zilibotti (2003). Democratic Public Good Provision. CEPR working paper No. 4044.

Hassler, J., K. Storesletten, and F. Zilibotti (2007). Democratic Public Good Provision. *Journal of Economic Theory 133*(1), 127–51.

Helpman, E., O. Isthoki, and S. Redding (2010). Inequality and Unemployment in a Global Economy. *Econometrica 78*(4), 1239–1283.

Irwin, D. (2017). *Clashing Over Commerce: A History of US Trade Policy*. Chicago University Press.

Jain, S. and S. Mukand (2003). Redistributive Promises and the Adoption of Economic Reform. *American Economic Review 93*(1), 256–264.

Johnson, H. G. (1951-1952). Optimum Welfare and Maximum Revenue Tariffs. *Review of Economic Studies 19*(1), 28–35.

Krusell, P. and J. Ríos-Rull (1996). Vested Interestes in a Positive Theory of Stagnation and Growth. *Review of Economic Studies 62*, 301–29.

Matschke, X. and S. M. Sherlund (2006). Do labor issues matter in the determination of u.s. trade policy? an empirical reevaluation. *American Economic Review 96*, 405–421.

Matsuyama, K. (1992). A Simple Model of Sectoral Adjustment. *Review of Economic Studies 59*, 375–87.

Mayer, W. (1984). Endogenous Tariff Formation. *American Economic Review 74*(5), 970–985.

Persson, T. and G. Tabellini (1994, June). Is Inequality Harmful for Growth? *American Economic Review 84*(3), 600–621.

Piketty, T. (2018). Brahmin Left vs Merchant Right: Rising Inequality and the Changing Structure of Political Conflict. Technical report, World Inequality Lab.

Razin, A., E. Sadka, and B. Suwankiri (2011). *Migration and the Welfare State*. Cambridge, MA: MIT Press.

Rodrik, D. (1994). The Rush to Free Trade. In S. Haggard and S. Webb (Eds.), *Voting for Reform: Democracy, Political Liberalization, and Economic Adjustment*, A world bank book 3, pp. 61–88. Oxford University Press.

Staiger, R. and G. Tabellini (1987). Discretionary Trade Policy and Excessive Protection. *American Economic Review 77*, 823–37.

Storesletten, K. (2000). Sustaining Fiscal Policy Through Immigration. *Journal of Political Economy 108*, 300–23.

Willmann, G. (2004). Pareto Gains from Trade: A Dynamic Counter-example. *Economics Letters 83*, 199–204.

A Proofs of Lemmas

A.1 Proof of Lemma 1

Proof. Totally differentiating the first order condition (7), Assumption 1, and the properties of the indirect utility function yields the required results:

$$\begin{aligned}
\frac{de(a;\tau_t,\tau_{t+1})}{da} &= -\frac{h_{ea}}{h_{ee}} > 0, \\
\frac{de(a;\tau_t,\tau_{t+1})}{d\beta} &= -\frac{h_e}{\beta h_{ee}} > 0. \\
\frac{de(a;\tau_t,\tau_{t+1})}{dp_t} &= \frac{v_p(p_t)/v(p_{t+1})}{\beta b h_{ee} p_{t+1}} > 0, \\
\frac{de(a;\tau_t,\tau_{t+1})}{dp_{t+1}} &= \frac{h_e(\alpha-1)}{h_{ee} p_{t+1}} > 0,
\end{aligned}$$

where we use $\alpha \equiv \frac{-v_p(p_{t+1})p_{t+1}}{v(p_{t+1})}$, by Roy's identity. □

A.2 Proof of Lemma 2

Proof. The first part of the proof follows directly from the first order condition in (15): evaluated at $t = 0$, $V_\tau \geq 0$ (implying a positive tariff) if and only if $\Delta \leq 0$. Claim 1 in Appendix C establishes that the second order condition, $V_{\tau\tau} < 0$ holds with strict equality.

To establish the signs on the derivatives, we first use that $V_{\tau a}(a, e)$ and $V_{\tau e}(a, e)$ are negative, as established in Claims 2-3 in Appendix C. Taking the total derivative of the first order condition in (15) with respect to a and τ, we have that $\frac{\partial \tau_t}{\partial a} = -\frac{V_{\tau a}}{V_{\tau\tau}} < 0$. Likewise, the total derivative of (15) with respect to e and τ yields $\frac{\partial \tau_t}{\partial e_{t-1}(a)} = -\frac{V_{\tau e}}{V_{\tau\tau}} < 0$. Next, $\frac{d\tau_t^o}{da} = \frac{\partial \tau_t}{\partial a} + \frac{\partial \tau_t}{\partial e(a)} \frac{\partial e(a)}{\partial a} < 0$, since the first two derivatives are negative (above) and the last factor is positive by Lemma 1. Finally, the total derivative of (15) with respect to Δ and τ yields $\frac{d\tau_t^o(\cdot)}{d\Delta_t} = -\frac{V_{\tau\Delta}}{V_{\tau\tau}} < 0$, since $V_{\tau\Delta} = -\frac{v_I p_t}{\tau_T} < 0$ and$V_{\tau\tau} < 0$ (above). □

A.3 Proof of Lemma 3

Proof. We begin by showing that the fixed point steady state solution $(\tilde{\Delta^m}, \tilde{\tau})$ is stable and unique as long if $\forall \Delta^m$, the steady state tariff schedule $\tau(\Delta^m)$ is not decreasing faster in Δ^m than is the schedule $\Delta^m(\tau)$ i.e.

$$\left.\frac{d\Delta^m}{d\tau}\right|_{\Delta^{m\,o}(\tau)} > \left.\frac{d\Delta^m}{d\tau}\right|_{\tau(\Delta^m)} \tag{A.1}$$

Notice that if $\Delta^{m\prime}(\tau) > 0$, this condition is assured, since we already have that $\tau'(\Delta^m) < 0$ from Lemma 2. But since we also allow for $\Delta^{m\prime}(\tau) < 0$, we need to make the preceding assumption to govern the relative slopes of the tariff and Δ^m functions.

Taking the derivative of the first order condition of the optimal tariff problem in (C.1) with respect to Δ^m, we have:

$$\begin{aligned} V_{\tau\tau} d\tau + V_{\tau\Delta^m} d\Delta^m &= 0, \text{which implies:} \\ \frac{d\Delta^m}{d\tau}\Big|_{\tau(\Delta^m)} &= -\frac{V_{\tau\tau}}{V_{\tau\Delta^m}} < 0. \end{aligned} \tag{A.2}$$

At the same time,

$$\frac{d\Delta^m}{d\tau} = \frac{d\Delta^m}{d\tau}\frac{dp}{d\tau} = \frac{d\Delta^m(p)}{dp}\left(\frac{-p^w}{\tau^2}\right) \tag{A.3}$$

Substituting (C.6) and (C.7) into (A.1) yields the condition in Assumption 2.

$$\begin{aligned} \frac{d\Delta^m(p)}{dp}\left(\frac{-p^w}{\tau^2}\right) &> -\frac{V_{\tau\tau}}{V_{\tau\Delta^m}}\Big|_{\tau^0} \\ \Leftrightarrow \frac{d\Delta^m(p)}{dp} &< \frac{\tau^2}{p^w}\frac{V_{\tau\tau}}{V_{\tau\Delta^m}}. \end{aligned}$$

Or rewritten in terms of parameters:

$$\left(\frac{h_e^2}{|h_{ee}|}\Big|_{a^m} - \int_a \frac{h_e^2}{|h_{ee}|} f(a)da\right) < \frac{\tau^2}{p^w}\left(\bar{h} + \frac{1}{bp^w}\right)(\alpha(\tau - (\tau - 1)\alpha)).$$

Thus, under Assumption 2, the steady state is stable and unique.

Furthermore, by Claim 7 below, if $\Delta^m(\tau)$ is increasing in τ, then the out-of-steady-state schedule, $\Delta^m(\tau_{t-1}, \tau_t)$ will also be increasing in τ_t, which ensures that the out-of-steady-state equilibrium is (also) stable and unique in that case. If instead $\Delta^m(\tau)$ is decreasing in τ faster than the tariff schedule, so that steady state is unique and stable, then Claim 7 implies that the same must also be true out of steady state: $\Delta^m(\tau_{t-1}, \tau_t)$ will be decreasing even faster in τ_t than the tariff schedule $\tau({\Delta^m}_t)$, which establishes stability outside of steady state as well. □

A.4 Proof of Lemma 4

Proof. Part (i): Totally differentiating the first order condition of the optimal tariff function in (15) with respect to τ and p^w yields $\frac{d\tau}{dp^w} = -\frac{V_{\tau p^w}}{V_{\tau\tau}}$. As already established, $V_{\tau\tau} < 0$ by the second order condition of the optimal tariff problem (Claim 1 in the appendix). In Claim 2 of the appendix, we show that $V_{\tau p^w} > 0 (\leq 0)$ if and only if $\Delta^m < 0$ (≥ 0), which yields the result. For part (ii), it is sufficient to show that evaluated at the fully compensating tariff, $V_\tau(\Delta^m; p^{w1})$ is strictly less than (greater than) zero if $\Delta^m < (>) 0$, which implies that the median voter would prefer a strictly smaller tariff (or, if $\Delta^m > 0$, a smaller import subsidy). Evaluating the first order condition of the optimal tariff problem at the new terms of trade and τ^{FC}, we have $V_\tau(\Delta^m; p^{w1})\Big|_{\tau^{FC}} = v_I \frac{p^1}{\tau^{FC}}\big(- \Delta^m + t^{FC}\tau^{FC}\frac{dE^{FC}}{d\tau}\big)$. For any given Δ^m, the initially optimal tariff, $\tau(\Delta^m)$ is given by the first order condition $\Delta^m = t^o\tau\frac{dE^o}{d\tau}$. Substituting in, and using that τ^{FC} holds the domestic price fixed at the initial level by definition ($p^1 = p^o$), we have: $V_\tau(\Delta^m; p^{w1}) = v_I \frac{p^o}{\tau^{FC}}\big(- t^o\tau\frac{dE^o}{d\tau} + t^{FC}\tau^{FC}\frac{dE^{FC}}{d\tau}\big)$. In the appendix (Claim 5), we prove that $t\tau\frac{dE}{d\tau}0$ is decreasing in the tariff level, which establishes the result: $V_\tau(\Delta^m; p^{w1})\Big|_{\tau^{FC}} < 0 \Rightarrow \tau(\Delta^m; p^{w1}) < \tau^{FC}$. □

B Proofs of Propositions

B.1 Proof of Proposition 1

Proof. To prove the first part of the proposition, it is sufficient to show that (a) the value of new steady state optimal tariff function, evaluated at the initial steady state value of $\tilde{\Delta}^{m0}$, is strictly less than fully compensating; and (b) the value of the new steady state Δ^m function, evaluated at the fully compensating tariff, coincides with the original steady state $\tilde{\Delta}^{m0}$. Part (a) follows directly from Lemma 4. Part (b) follows immediately from the definition of $\Delta^m = (1-\alpha)b[h(a^m, p) - \bar{h}(p)]$, which is independent of τ, holding p fixed. Since by definition τ^{FC} would hold the domestic price unchanged, $\Delta^m(\tau^{FC}; p^{w1}) = \Delta^m(\tilde{p}^0) = \tilde{\Delta}^{m0}$. Together with the (assumed) regularity conditions over $h(\cdot)$ to assure a stable steady state, (a) and (b) establish Part 1 of the Proposition. Part 2 of the proposition follows directly. Since the new steady state tariff is less than fully compensating, the new domestic price must be strictly higher in the new steady state ($\tilde{p}^1 > \tilde{p}^0$). Education is monotonic in the domestic price, and therefore the new steady state education level will be higher than the

initial steady state education level for all individuals. □

B.2 Proof of Proposition 2

Proof. Applying Lemma 4 at the initial steady state value of ($\tilde{\Delta}^{m0}$) establishes the result. □

B.3 Proof of Proposition 3

Proof. Proposition 2, establishes point (i) directly. We also use it to prove part (ii). At the time of the shock, $\Delta^m_T = \tilde{\Delta}^{m0}$ and $\tau_T = \tau(\tilde{\Delta}^{m0}; p^{w1}) > \tau(\tilde{\Delta}^{m0}; p^{w0})$. As established as part of the proof for Proposition 2, the new steady state $\Delta^m(\tau; p^{w1})$ locus takes a value of $\tilde{\Delta}^{m0}$ at $\tau_{FC} > \tau_T$. Under this case's assumption that $\frac{d\Delta^m(p)}{dp} > 0 > \frac{d\Delta^m}{d\tau}$, this schedule is decreasing in τ. Since $\tau(\Delta^m; p^{w1})$ is also (always) decreasing in Δ^m, the new steady state $\Delta^m(\tau)$ and $\tau(\Delta^m)$ schedules must intersect at some value where $\Delta^m > \tilde{\Delta}^{m0} = \Delta^m_T$ and $\tau(\overset{\approx}{\Delta^m}; p^{w1}) < \tau_T$.

To establish part (iii) of the proposition, we use induction to trace out the fixed point equilibrium values of τ and Δ^m in successive periods after T. Beginning with period $T+1$, consider a candidate value of $\tau_{T+1} = \tau_T$, which would imply that $\Delta^m_{T+1} = \Delta^m(p_T, p_T) \equiv \Delta^m(p_T)$. Note that this candidate $\Delta^m(p_T) > \Delta^m_T = \tilde{\Delta}^{m0} = \Delta^m(\tilde{p}^0)$, since $p_T > \tilde{p}^0$ (by Proposition 2) and $\Delta^{m1}(p) > 0$ by assumption. This candidate cannot be a steady state, however, since then we would have $\tau_{T+1} = \tau(\Delta^m(p_T); p^{w1}) < \tau_T = \tau(\Delta^m(\tilde{p}^0); p^{w1})$, resulting in a contradiction. Now let $\Delta^m_{T+1} = \Delta^m(p^{w1}/\tau_T, p^{w1}/\tau_{T+1})$, where $\tau_{T+1} = \tau(\Delta^m_{T+1}; p^{w1})$. Compared to our benchmark, it must be that $\tau_{T+1} < \tau_T$ and $\Delta^m_{T+1} > \Delta^m(p^{w1}/\tau_T, p^{w1}/\tau_T) = \Delta^m(p_T) > \Delta^m_T$, since, according to our regularity conditions, $\Delta^m(\tau_{t-1}, \tau_t)$ decreases faster in its second argument than $\tau(\Delta^m_t; p^{w1})$ decreases in Δ^m. This argument can be repeated for every subsequent period, establishing that transition to the new steady state is a monotonic decline in tariffs. The rest is immediate, as the tariff falls, the domestic price rises, and so – by Lemma 1 – education rises for all workers. □

B.4 Proof of Proposition 4

Proof. As in the previous proof, point (i) is established directly by Proposition 2, which we also use to prove part (ii). Here again, we use that the new steady state locus $\Delta^m(\tau; p^{w1})$ takes the value of $\tilde{\Delta}^{m0}$ evaluated at τ_{FC}, whereas the new steady state tariff locus evaluated at $\tilde{\Delta}^{m0}$ takes a strictly smaller value: i.e. $\tau(\tilde{\Delta}^{m0}; p^{w1}) < \tau^{FC}$. Under the assumption that $\frac{d\Delta^m(p)}{dp} < 0$, the steady state $\Delta^m(\tau; p^{w1})$ schedule is *increasing* in τ. Since $\tau(\Delta^m; p^{w1})$ is always decreasing in Δ^m, this implies that the steady state $\Delta^m(\tau)$ and $\tau(\Delta^m)$ schedules must intersect at some value $\tilde{\Delta}^{m1} < \tilde{\Delta}^{m0}$ and $\tilde{\tau}^1 > \tau_T > \tilde{\tau}^0$. In the razor's edge case in which $\frac{d\Delta^m(p)}{dp} = 0$, transition will be instantaneous at time T and $\tilde{\Delta}^{m1} = \tilde{\Delta}^{m0}$ and $\tilde{\tau}^1 = \tau_T > \tilde{\tau}^0$.

To establish part (iii) of the proposition, Δ^m and τ (and hence p) oscillate around the new steady state, and converge toward it. To establish the oscillation, we consider successive periods subsequent to T, beginning with period $T+1$. For period $T+1$, we show that $\tau_{T+1} \geq \tau_T$ using proof by contradiction. Suppose not, s.t. $\tau_{T+1} < \tau_T$ and thus $p_{T+1} > p_T$. Since Δ^m is decreasing in p by assumption, this would imply $\Delta^m_{T+1} \equiv \Delta^m(p_T, p_{T+1}) < \Delta^m(p_T, p_T) \equiv \Delta^m(p_T) < \tilde{\Delta}^{m0}$. But since the tariff schedule $\tau(\Delta^m; p^w)$ is decreasing in Δ^m, this would then imply that $\tau(\Delta^m_{T+1}; p^{w1}) > \tau(\tilde{\Delta}^{m0}; p^{w1}) = \tau_T$, which is a contradiction. It must also be true that $\tau_{T+1} \geq \tilde{\tau}^1$. Again, suppose not: i.e. let $\tau_{T+1} < \tilde{\tau}^1$, which would imply that $p_{T+1} > \tilde{p}^1$. Since $p_T > \tilde{p}^1$ from part $ii)$ of the proposition, it would then be the case that $\Delta^m(p_T, p_{T+1}) < \tilde{\Delta}^{m1}$, which would in turn imply that $\tau_{T+1} > \tilde{\tau}^1$: another contradiction. Thus, $\tau_{T+1} \geq \tilde{\tau}^1 \geq \tau_T$ and $\Delta^m_{T+1} > \tilde{\Delta}^{m1} > \tilde{\Delta}^{m0}$

We can follow the same procedure to show that $\tau_{T+2} \leq \tilde{\tau}^1 \leq \tau_{T+1}$. Suppose not. This would then imply that both $\tau_{T+1}, \tau_{T+2} > \tilde{\tau}^1$, so that $p_{T+2}, p_{T+1} < \tilde{p}^1$. But then, $\Delta^m_{T+2} \equiv \Delta^m(p_{T+1}, p_{T+2}) < \tilde{\Delta}^{m1}$. And since the tariff is decreasing in Δ^m, this would mean that $\tau(\Delta^m_{T+2}; p^{w1}) > \tilde{\tau}^1$: contradiction. Thus, it must be true that $\tau_{T+2} \leq \tilde{\tau}^1 \leq \tau_{T+1}$ and likewise $\Delta^m_{T+2} \leq \tilde{\tau}^1 \leq \tau_{T+1}$ It is obvious that this same proof by contradiction will hold for all subsequent periods $T+t$ where $t \geq 2$. To show convergence, we need only establish that after each full oscillation, the state and policy outcome will be closer to the new steady state than before. Start at (Δ^m_T, τ_T). Next consider $(\Delta^m(p_T, p_T), \tau(\Delta^m(p_T, p_T))$ which lies on the new τ-schedule, but farther from the new steady state than the actual $(\Delta^m_{T+1}, \tau_{T+1})$ from above, because the out-of-steady-state Δ^m-schedule has a finite partial derivative in its second argument (by Claim 7). Repeat this argument for $T+2$ in the opposite direction: because the slope of the new τ-schedule is less than the slope of the new steady state Δ^m-schedule, it must hold that $(\Delta^m_{T+2}, \tau_{T+2})$ is closer to the new steady state than (Δ^m_T, τ_T). We can repeat this argument for all successive full oscillations, which establishes the result.

□

B.5 Proof of Proposition 5

Proof. The first part of the proposition is established by showing that an increase in b causes the tariff to rise immediately at time T. Note first that holding education fixed, an increase in b magnifies initial inequality: $\Delta^m_T = (1-\alpha)b'[h(a^m, \tilde{e}^0(a) - \bar{h}(a, \tilde{e}^0(a))f(a)da] = \frac{b'}{b^0}\tilde{\Delta}^{m0}$. Since $\tilde{\Delta}^{m0} < 0$ it must be that $\Delta^m_T < \tilde{\Delta}^{m0}$. All else equal, this would increase the tariff. But the tariff locus also shifts, so to establish the net effect, we need to show that holding education fixed, $\frac{d\tau}{db} > 0$. Taking the total derivative of the first order condition of the optimal tariff problem, yields $V_{\tau\tau}d\tau + V_{\tau b}db = 0$, or, $\left.\frac{d\tau}{db}\right|_{\tau(\Delta^m)} = -\frac{V_{\tau b}}{V_{\tau\tau}}$. Recall from Claim 1 that $\left.V_{\tau\tau}\right|_{\tau(\Delta^m)} < 0$. Thus, the sign of $\frac{d\tau}{db}$ is given by the sign of $\left.V_{\tau b}(\tau, b)\right|_{\tau(\Delta^m)}$. Claim 6 in the Appendix proves that, holding education fixed, $\left.V_{\tau b}\right|_{\tau(\Delta^m)} > 0 \iff \tilde{\Delta}^{m0} < 0$, which establishes part (i) of the proposition.

Subsequent to time T, there are two possibilities depending on the sign of $\frac{d\Delta^m}{dp}$. First,

consider the "overshooting" case (a), in which $\frac{d\Delta^m}{dp} \geq 0$. We need only to show that $\Delta^m_{T+1} \geq \Delta^m_T$, after which the transition proceeds via monotonic tariff adjustment just as in Proposition 3. (As before, under the razor's edge case in which $\frac{d\Delta^m}{dp} = 0$, transition will be immediate.) Using the definition of $\Delta^m(\cdot)$ we first show that $\Delta^m(\tau_T, \tau_T; b') \geq \Delta^m_T$: $\Delta^m(\tau_T, \tau_T; b') = \Delta^m_T + (1-\alpha)b'\left[\left(\frac{dh(a^m)}{db}\right) - \left(\frac{d\bar{h}}{db}\right)\right] = \Delta^m_T + (1-\alpha)\left[\left(\frac{h_e^2}{-h_{ee}}\right)\Big|_{a^m} - \int_a \left(\frac{h_e^2}{-h_{ee}}\right) f(a)da\right]$. But $\frac{d\Delta^m}{dp} \geq 0$ implies that $\left(\frac{h_e^2}{-h_{ee}}\right)\Big|_{a^m} \geq \int_a \left(\frac{h_e^2}{-h_{ee}}\right) f(a)da$. Thus, it must hold that $\Delta^m(\tau_T, \tau_T; b') \geq \Delta^m_T$. Then, since $\tau(\Delta^m_t)$ is also decreasing in Δ^m, the fixed point intersection of $\tau(\Delta^m_{T+1})$ and $\Delta^m(\tau_T, \tau_{T+1})$ must occur for some $\Delta^m_{T+1} \geq \Delta^m_T$ and $\tau_{T+1} \leq \tau_T$.

We use a similar technique for Case (b), in which $\frac{d\Delta^m}{dp} < 0$. Again, we need to show only that $\Delta^m_{T+1} < \Delta^m_T$, after which transition will proceed via the same oscillating tariff pattern in Proposition 4. Applying the same logic as above, we have $\Delta^m(\tau_T, \tau_T; b') = \Delta^m_T + (1-\alpha)\left[\left(\frac{h_e^2}{-h_{ee}}\right)\Big|_{a^m} - \int_a \left(\frac{h_e^2}{-h_{ee}}\right) f(a)da\right]$. But when $\frac{d\Delta^m}{dp} < 0$ the second term is negative, so that $\Delta^m(\tau_T, \tau_T; b') < \Delta^m_T$. Finally, when $\frac{d\Delta^m}{dp} < 0$, $\Delta^m(\tau_T, \tau_{T+1})$ is increasing in the second argument, which implies that $\Delta^m_{T+1} < \Delta^m_T$ and $\tau_{T+1} > \tau_T$. □

C Proofs of Claims

The following claims establish a series of useful properties that we use in the proofs above.

Claim 1. $V_{\tau\tau}(\tau_t, \Delta^m_t)\Big|_{\tau(\Delta^m_t)} < 0$

Proof. Recall the expression for V_τ from the first order condition in (15):

$$V_\tau = \nu(p_t)p_t\left[-\frac{\Delta^m_t}{\tau_t} + (\tau_t - 1)\frac{d\bar{E}_t}{d\tau_t}\right]. \tag{C.1}$$

Taking the derivative with respect to τ_t:

$$V_{\tau\tau} = \nu(p_t)p_t\left(\frac{\Delta^m_t}{\tau_t^2} + \frac{d\bar{E}_t}{d\tau_t} + (\tau_t - 1)\frac{d^2\bar{E}_t}{d\tau_t^2}\right),$$

where the second line uses the envelope condition and we use $\nu(p_t) = \nu_I$. Evaluated at the optimal tariff: $\frac{\Delta^m_t}{\tau_t} = (\tau_t - 1)\frac{d\bar{E}_t}{d\tau_t}$ from the first order condition,

$$V_{\tau\tau}\Big|_{\tau(\Delta^m_t)} = \nu(p_t)p_t\left[\left(\frac{2\tau_t - 1}{\tau_t}\right)\frac{d\bar{E}_t}{d\tau_t} + (\tau_t - 1)\frac{d^2\bar{E}_t}{d\tau_t^2}\right]. \tag{C.2}$$

The expression for total net exports is given implicitly by $\bar{E}_t = \bar{x}^s_t - \bar{d}^s_t$, or, with Cobb-Douglass preferences:

$$\bar{E}_t = \bar{x}^s_t - \frac{\alpha}{p_t}(p_t\bar{x}^s_t + 1 + (\tau_t - 1)p_t\bar{E}_t),$$

which can be rearranged to yield the expression for total net exports:

$$\bar{E}_t = \bar{E}(\tau_t, \mathbf{e}_{t-1}; p_t^w) = \frac{(1-\alpha)\bar{x}_t^s(\mathbf{e}_{t-1}) - \frac{\tau_t \alpha}{p_t^w}}{(1+(\tau_t-1)\alpha)}. \tag{C.3}$$

Taking the derivative with respect to τ_t:

$$\begin{aligned}
\frac{d\bar{E}_t}{d\tau_t} &= \frac{-\alpha/p_t^w}{(1+(\tau_t-1)\alpha)} - \alpha\left(\frac{(1-\alpha)\bar{x}_t^s - \tau_t\alpha/p_t^w}{(1+(\tau_t-1)\alpha)^2}\right) \\
&= \left(\frac{-\alpha/p_t^w}{(1+(\tau_t-1)\alpha)^2}\right)((1-\alpha)\bar{x}_t^s p_t^w - \tau_t\alpha + 1 + \alpha(\tau_t-1)) \\
&= \left(\frac{-\alpha(1-\alpha)}{(1+(\tau_t-1)\alpha)^2}\right)\left(\bar{x}_t^s + \frac{1}{p_t^w}\right) < 0. \qquad \text{(C.4)}
\end{aligned}$$

Taking the derivative again,

$$\begin{aligned}
\frac{d^2E_t}{d\tau_t^2} &= \underbrace{\left(\frac{-\alpha(1-\alpha)}{(1+(\tau_t-1)\alpha)^2}\right)\left(\bar{x}_t^s + \frac{1}{p_t^w}\right)}_{=\frac{dE_t}{d\tau_t}(<0)}\left(\frac{-2\alpha}{1+(\tau_t-1)\alpha}\right) \\
&= \frac{dE_t}{d\tau_t}\left(\frac{-2\alpha}{1+(\tau_t-1)\alpha}\right) > 0 \qquad \text{(C.5)}
\end{aligned}$$

Substituting (C.4) and (C.5) into (C.2) yields:

$$\begin{aligned}
V_{\tau\tau}\Big|_{\tau(\Delta_t^m)} &= \nu(p_t)p_t\left(\frac{2\tau_t-1}{\tau_t} - \frac{2(\tau_t-1)\alpha}{(1+(\tau_t-1)\alpha)}\right)\frac{dE_t}{d\tau_t} \\
&= \underbrace{\nu(p_t)p_t}_{+}\left[\underbrace{\left(\frac{1+(\tau_t-1)(1-\alpha)}{\tau_t(1+(\tau_t-1)\alpha)}\right)}_{+}\underbrace{\frac{dE_t}{d\tau_t}}_{-}\right] < 0.
\end{aligned}$$

□

Claim 2. $V_{\tau p^w}(\Delta_t^m, p_t^w)\Big|_{\tau(\Delta_t^m)} > 0$ *if and only if* $\Delta_t^m < 0$.

Proof. As above, start by taking the derivative of (C.1), now with respect to p_t^w, and use the envelope condition:

$$V_{\tau p^w}\Big|_{\tau} = \nu(p_t)p_t\Big((\tau_t-1)\frac{\partial^2\bar{E}_t}{\partial\tau_t\partial p_t^w}\Big). \tag{C.6}$$

Then take the derivative of (C.4) with respect to p_t^w to get:

$$\frac{\partial^2\bar{E}_t}{\partial\tau\partial p_t^w} = \frac{\alpha(1-\alpha)}{(1-(\tau_t-1)\alpha)^2}\frac{1}{p_t^{w\,2}} > 0 \tag{C.7}$$

Substituting (C.7) into (C.6) yields:

$$V_{\tau p^w} = \underbrace{\nu(p_t)p_t}_{+}\Big((\tau_t - 1)\underbrace{\frac{dE_t^2}{d\tau_t dp_t^w}}_{+}\Big).$$

Thus, the sign depends on the sign of the initially optimal tariff, whether $\tau(\Delta_t^m)$ is greater or less than one, which depends in turn on the sign of Δ_t^m. Summarizing:

$$\Delta_t^m < 0, \tau(\Delta_t^m) > 1 \;\Rightarrow\; V_{\tau p^w} > 0$$
$$\Delta_t^m \geq 0;\; \tau(\Delta_t^m) \leq 1 \;\Rightarrow\; V_{\tau p^w} \leq 0.$$

□

Claim 3. $V_{\tau a}(a, \mathbf{e}_{t-1})\Big|_{\tau(\Delta_t^m)} < 0.$

Proof. Taking the derivative of (C.1) with respect to a (holding $\mathbf{e}_{t-1}$ fixed), we have:

$$V_{\tau a} = \nu(p_t)p_t\left(-\frac{1}{\tau_t}\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial a}\right)$$

Recall, that $\Delta(a, \mathbf{e}_{t-1}) = (1-\alpha)b[h(a, e_{t-1}(a)) - \bar{h}(\mathbf{e}_{t-1})]$, which implies:

$$\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial a} = (1-\alpha)b\frac{\partial h(a, e_{t-1}(a))}{\partial a} > 0$$

Substituting gives the result:

$$V_{\tau a}\Big|_{\tau=\tau(\Delta_t^m)} = \underbrace{\nu(p_t)p_t}_{+}\left(-\frac{1}{\tau_t}\right)\underbrace{\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial a}}_{+} < 0$$

□

Claim 4. $V_{\tau e(a)}(a, \mathbf{e}_{t-1})\Big|_{\tau(\Delta_t^m)} < 0$

Proof. Taking the derivative of (C.1) with respect to $e_{t-1}(a)$ (holding a the remaining distribution of education fixed), we have:

$$V_{\tau e(a)} = \nu(p_t)p_t\left(-\frac{1}{\tau_t}\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial e(a)}\right)$$

where,

$$\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial e(a)} = (1-\alpha)b\frac{\partial h(a, e_{t-1}(a))}{\partial e(a)} > 0.$$

Thus,

$$V_{\tau e}\Big|_{\tau(\Delta_t^m)} = \underbrace{\nu(p_t)p_t}_{+}\left(-\frac{1}{\tau_t}\right)\underbrace{\left(\frac{\partial\Delta(a, \mathbf{e}_{t-1})}{\partial e(a)}\right)}_{+} < 0$$

□

Claim 5. *The expression* $(\tau_t - 1)\tau_t \frac{dE_t}{d\tau_t}$ *is decreasing in* τ_t.

Proof.

$$\frac{d}{d\tau_t}\left((\tau_t - 1)\tau_t \frac{dE_t}{d\tau_t}\right) = \frac{dE_t}{d\tau_t}(2\tau_t - 1) + (\tau_t - 1)\tau_t \frac{d^2 E_t}{d\tau_t}.$$

Substituting from (C.5):

$$\begin{aligned}\frac{d}{d\tau_t}\left((\tau_t - 1)\tau_t \frac{dE_t}{d\tau_t}\right) &= \frac{dE_t}{d\tau_t}\left(2\tau_t - 1 - \frac{2\alpha(\tau_t - 1)\tau_t}{(1 + (\tau_t - 1)\alpha)}\right). \\ &= \underbrace{\frac{dE_t}{d\tau_t}}_{-}\underbrace{\left(\frac{\tau_t + (\tau_t - 1)(1 - \alpha)}{1 + (\tau_t - 1)\alpha}\right)}_{+} < 0.\end{aligned}$$

□

Claim 6. *Holding the distribution of education fixed,* $V_{\tau b}|_{\tau(\Delta_t^m)} > 0$ *if and only if* $\Delta_t^m < 0$.

Proof. Taking the derivative of the first order condition in (C.1) with respect to b, holding education levels fixed at $\mathbf{e}_{t-1}$:

$$V_{\tau b} = \nu_I \frac{p_t}{\tau_t}\left\{(\tau_t - 1)\tau_t \frac{\partial^2 \bar{E}_t}{\partial \tau \partial b} - \frac{\partial \Delta_t^m(\mathbf{e}_{t-1}, b)}{\partial b}\right\}. \tag{C.8}$$

Using $h_t(a^m) \equiv h(a^m, e_{t-1}(a^m))$ and $\bar{h}_t \equiv \int_a h(a, e_{t-1}(a))f(a)da$ as shorthand, recall the definition of Δ_t^m:

$$\Delta_t^m = (1 - \alpha)b(h_t(a^m) - \bar{h}_t),$$

which implies in turn that:

$$\frac{\partial \Delta^m(\mathbf{e}_{t-1}, b)}{\partial b} = (1 - \alpha)(h_t(a^m) - \bar{h}_t) = \frac{\Delta_t^{m0}}{b} = (\tau_t - 1)\tau_t\left(\frac{1}{b}\right)\frac{d\bar{E}_t}{d\tau} \tag{C.9}$$

where Δ_t^{m0} denotes the initial value of Δ_t^m and the last equality uses the first order condition for the initially optimal tariff.

Substituting (C.9) into (C.8) and collecting terms, we have:

$$V_{\tau b} = \nu_I p_t(\tau_t - 1)\left\{\frac{\partial^2 \bar{E}_t}{\partial \tau \partial b} - \frac{1}{b}\frac{d\bar{E}_t}{d\tau_t}\right\} \tag{C.10}$$

Taking the derivative of the expression in (C.4) with respect to b yeilds:

$$\frac{\partial^2 \bar{E}_t}{\partial \tau \partial b} = \frac{-\alpha(1 - \alpha)}{1 + (\tau_t - 1)\alpha}\bar{h} \tag{C.11}$$

Finally, substituting (C.4) and (C.11) into (C.10) delivers the result:

$$\begin{aligned} V_{\tau b} &= \nu_I p_t(\tau_t - 1)\left\{\frac{-\alpha(1-\alpha)}{(1+(\tau_t-1)\alpha)^2}\bar{h} - \left[\frac{-\alpha(1-\alpha)}{(1+(\tau_t-1)\alpha)^2}\left(\bar{h} + \frac{1}{p_t^w h}\right)\right]\right\} \\ &= \nu_I p_t(\tau_t - 1)\underbrace{\left\{\frac{\alpha(1-\alpha)}{(1+(\tau_t-1)\alpha)^2 p_t^w b}\right\}}_{+} > 0 \iff \tau(\Delta_t^m) > 1 \quad \text{which } \Leftrightarrow \Delta_t^m < 0. \end{aligned}$$

□

Claim 7. *The steady state function $\Delta^m(p)$ is more responsive to changes in p than is the out-of-steady-steady schedule $\Delta^m(p_t, p_{t+1})\forall t$ in either argument.*

1. *If* $\frac{d\Delta^m(p)}{dp} > 0$ *then* $\frac{d\Delta^m(p)}{dp} > \frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_t}$, $\frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_{t+1}} > 0$; *and*
2. *If* $\frac{d\Delta^m(p)}{dp} \leq 0$, *then* $\frac{d\Delta^m(p)}{dp} \leq \frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_t}$, $\frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_{t+1}} \leq 0$.

In other words, if the steady state is unique and stable, then the same must be true for the out of state equilibrium pairs, $(\Delta^m{}_t, \tau_t)\forall t$.

Proof. From the definition in (15):

$$\Delta^m(p_t, p_{t+1}) = (1-\alpha)b\left[h(a^m, e(a^m, p_t, p_{t+1})) - \int_a h(a, e(a, p_t, p_{t+1}))f(a)da\right]$$

So,

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_t} = (1-\alpha)b\left[h_e\frac{\partial e(a^m; p_t, p_{t+1})}{\partial p_t} - \int h_e\frac{\partial e(a; p_t, p_{t+1})}{\partial p_t}f(a)da\right] \quad \text{(C.12)}$$

Likewise,

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_{t+1}} = (1-\alpha)b\left[h_e\frac{\partial e(a^m; p_t, p_{t+1})}{\partial p_{t+1}} - \int h_e\frac{\partial e(a; p_t, p_{t+1})}{\partial p_t}f(a)da\right] \quad \text{(C.13)}$$

From Lemma 1:

$$\frac{\partial e(a; p_t, p_{t+1})}{\partial p_t} = \frac{\alpha}{1-\alpha}\,\frac{p_{t+1}}{p_t}\,\frac{\partial e(a; p_t, p_{t+1})}{\partial p_{t+1}} \quad \text{(C.14)}$$

Combining (C.12)-(C.14) yields

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_t} = \frac{\alpha}{1-\alpha}\,\frac{p_{t+1}}{p_t}\,\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_{t+1}}$$

Thus,

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_t} > 0 \quad \iff \quad \frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_{t+1}} > 0.$$

Finally, from the definition of the steady state schedule $\Delta^m(p)$:

$$\frac{d\Delta^m(p)}{dp} \equiv \left(\frac{\partial \Delta^m(p_t, p_{t+1})}{\partial p_t} + \frac{\partial \Delta^m(p_t, p_{t+1})}{\partial p_{t+1}} \right) \Bigg|_{p_t = p_{t+1} = p}$$

So if

$$\frac{d\Delta^m(p)}{dp} > 0 \quad (\leq 0),$$
$$\frac{\partial \Delta^m(p_t, p_{t+1})}{\partial p_t} \quad \text{and} \quad \frac{\partial \Delta^m(p_t, p_{t+1})}{\partial p_{t+1}} > 0 \ (\leq 0).$$

□