24 DEC

THE TWENTY-FOURTH DUBROVNIK ECONOMIC CONFERENCE

Organized by the Croatian National Bank

Emily Blanchard and Gerald Willmann

# Unequal Gains, Prolonged Pain: A Model of Protectionist Overshooting and Escalation

Hotel "Grand Villa Argentina"
Dubrovnik
June 3 – 5, 2018

Draft version
Please do not quote

CROATIAN NATIONAL BANK

# Unequal Gains, Prolonged Pain: A Model of Protectionist Overshooting and Escalation*

Emily Blanchard† Gerald Willmann‡

May 24, 2018

**PRELIMINARY**

**Abstract**

This paper studies democratic political responses to macroeconomic shocks in the short and long run. We develop a model in which economic adjustment is slower than potential political change, and show that changes global marketplace can exacerbate political polarization, leading to surges in popular support for distortionary economic policies. Applied to the trade policy context, we find that when the returns to openness are concentrated at the top of the income distribution, an exogenous terms-of-trade improvement or skill-biased technological change will lead to a spike in protectionism that blunts the incentives of the younger generation to acquire education. In the long run, the initial spike in protectionism will gradually diminish if – and only if – education enables less-skilled workers to catch up with the aggregate economy. The more unequal the initial distribution of gains and losses from the shock among the population, the greater and longer lasting the induced protectionism: unequal gains, prolonged pain. Evidence on key data markers suggested by the model exhibits patterns consistent with the recent populist support for Brexit and Trump.

*JEL Classifications: F5, D7, E6*

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

---

*We are grateful our colleagues and seminar participants for extensive and constructive feedback.
†Tuck School of Business at Dartmouth and CEPR; emily.blanchard@tuck.dartmouth.edu
‡University of Bielefeld and IfW Kiel; gwillmann@wiwi.uni-bielefeld.de

Globalization has suffered a spate of sharp democratic rebukes over the past two years, including the UK 'Brexit' vote and the US presidential election of trade-skeptic Donald Trump. This surge in economic nationalism has confounded many of 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 in driving job losses as foreign competition. While these arguments are correct, they are incomplete. To understand the forces driving today's protectionist groundswell, we need to incorporate both dynamic frictions and static distribution into our workhorse understanding of political support for trade.

In this paper, we show that when economic adjustment is slow and the gains from trade are skewed toward the top, protectionist surges are a natural democratic response to unanticipated macroeconomic changes. Crucially, this prediction holds even when shocks deliver immediate aggregate welfare gains and even if those gains will eventually be shared by a majority of voters. The reason is an inherent timing mis-match: structural change takes time, while politics can respond more quickly. Even if in the long run most individuals will be 'winners' from more open borders, in the short run, many will lose because labor market frictions slow their potential to respond to a changing marketplace.

The core of our paper develops 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 then use the model to evaluate the extent to which domestic economic policies or multilateral trade disciplines will soften or sharpen the political consequences of macroeconomic shocks for trade policy. A short final section of the paper considers recent data from the US, UK, and comparable trading partners in the context of our theory.

The model features a small open economy with overlapping generations of heterogenous workers who make endogenous human capital investments. In each generation, young 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).[1] We consider permanent, unanticipated, 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 exercise generates three key insights. First, differential 'stickiness' between economic and political change can lead to policy volatility, including the potential for *protectionist overshooting* in response to an unanticipated terms-of-trade improvement or skill-biased technological change.[2] If politics can respond to shocks more quickly than labor markets can adjust, then even if the shock will eventually lead to lower tariffs, the short run response will be an increase in trade protection. This surge in protectionism then slows the subsequent process of political and economic adjustment by blunting the incentive for younger workers to acquire human capital. Inequality falls as the adversely-affected less-skilled workers 'catch up' to the rest of the economy, but the process takes time.

Second, the skewness of the initial 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 it is entirely possible that a shock will exacerbate the underlying skewness in the distribution of human capital, even after workers have had time to upgrade their skill sets. We show that if the shock induces a greater increase in human capital at the top of the distribution than at the bottom, the long run equilibrium will be characterized by *protectionist escalation*: after the initial protectionist surge, 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

[1]Under majoritarian voting, the key feature of the distribution is the difference in the tariff exposure between the politically pivotal median voter and the overall economy; under more general political systems, a different moment of the population distribution may drive formal results. The upshot remains the same, however: the overall distribution of gains and losses – not just the aggregate – is critical in determining policy. See Alesina and Rodrik (1994) for a compelling discussion of this point.

[2]We intentionally use the term *overshooting* to evoke Dornbusch (1976). Just as exchange rate overshooting in Dornbusch (1976) is generated by the marriage of sticky prices with the immediate response of market expectations, protectionist overshooting in our model is generated by the interplay between sticky labor markets and immediate political response.

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 can be caused by technology, not trade: even if automation is *entirely* responsible for today's increasing economic polarization, the political consequences for globalization may be the same. More broadly, anything that increases the dispersion in the distribution of the gains from trade can sharpen voters' incentives to tilt market wages in their favor. If tariffs are democratically determined, economic nationalism may be the inevitable and natural consequence.

We use the model to evaluate the extent to which protectionist surges are exacerbated or softened in the presence of other domestic economic policies like redistributive taxes and transfers or education spending. We show that unconditional lump sum redistribution (e.g. basic income) workers is unlikely to mitigate economic populism, even if the transfers reduce income inequality. As long as some part of workers' earnings are linked to domestic prices, voters will have an incentive to manipulate tariffs. By the same logic, progressive taxes that reduce inequality in (post-tax) earnings *can* lessen the magnitude and duration of protectionist surges, but at the risk of discouraging investment in education. In contrast, education policies 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 subsidies increase educational attainment disproportionately among those workers already at the top of the distribution, they will only worsen political polarization (and thus protectionism).[3] Finally, 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 exercise 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 use income inequality to proxy differences in 'unequal gains' across countries and over time. 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 mobility offers a rough indication the extent to which workers can overcome

[3]These findings echo Blanchard and Willmann (2016), applied to a dynamic setting with endogenous voting.

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.

Finally, before continuing, it is worth noting that although our model is tailored to understand the recent surge in protectionism, the basic theoretical insight is much broader. Economic adjustment could instead take the form of physical capital accumulation, changes in land use, or technology adoption. Likewise, the median voter rule acts as a tractable stand-in for nearly any political environment in which the underlying distribution of voters' preferences matters.[4] Our key insight is that 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 can be extended to a broad set of applications including immigration, political responses to climate change, tax reform, and beyond.

The paper proceeds as follows. The next section reviews the important and diverse related literature that precedes us, while Section 2 presents our model and characterizes economic and political steady states. Section 3 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 4, we use the model to shed light on domestic and multilateral policies that may exacerbate or mitigate populist protectionist surges. Section 5 presents data on empirical indicators suggested by the model and Section 6 concludes.

## 1 Related Literature

In this paper, we build on a broad foundation of existing work in trade, political economy, and macroeconomics. Our work is motivated in part by the important recent empirical work of Artuç, Chaudhuri, and McLaren (2010), Autor, Dorn, and Hanson (2013), Dix-Caneiro (2014), and others in highlighting the important role that adjustment costs play in shaping the distributional consequences of trade. Along a different dimension, recent empirical findings by Bown and Crowley (2012) and Hillberry and McCalman (2011) both suggest that flexible protectionist policy instruments (anti-dumping cases and other temporary trade

[4] Alesina and Rodrik (1994) make a compelling case for this point, and argue the relevance of median voter insights even in a dictatorship as long as there is some threat of a coup.

barriers) respond to global economic shocks in a manner consistent with the 'overshooting' mechanism we highlight in this paper.[5]

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.[6] At the same time, the political hysteresis in our model continues the tradition of Fernandez and Rodrik (1991), who demonstrate the potential for endogenous resistance to trade reform due to uncertainty. From a modeling perspective, our paper also echoes the "putty-clay" labor market structure in Matsuyama (1992).

Our work is also reminiscent of Staiger and Tabellini (1987), who highlight the importance of time consistency (and its absence) in driving "excessive" protection, which can occur if long-lived governments cannot pre-commit to future free trade.[7] While our overlapping generations framework is quite different (by definition, the democratically most-preferred tariff is not "excessive"), their broader point about the potential role for tariff commitments is also salient in our model, as we discuss in the context of multilateral escape clauses. Most closely, we build on our previous work on dynamic endogenous trade policy in Blanchard and Willmann (2011). Whereas the model in our earlier paper was limited to a binary policy choice, binary skill acquisition decision, and comparison of political steady states, we move beyond each of these limitations in the present paper, allowing us to study transition dynamics in a richer setting.

In the growth 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 an overlapping generations and slow adjustment, but none of these allow for both the differential speed of real versus political adjustment and the endogenous evolution of political preferences (e.g. via income), that together give rise to our overshooting mechanism.

More recently, Acemoğlu, Naidu, Restrepo, and Robinson (2015) highlight the interplay between democracy and redistribution and find empirical support for the importance

[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.

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

[7]Brainard and Verdier (1997) offer a complementary model driven by special interest politics, in which declining import-competing industries can slow their senscence via costly lobbying for protection.

of the politically pivotal middle class, particularly in promoting redistribution and structural change through secondary schooling.[8] Outside the political economy framework, but also closely related is the important work by Acemoğlu, Gancia, and Zilibotti (2015), who demonstrate the potential for increased openness via offshoring to drive skill-biased technical change, increasing inequality through complementary channels.

Finally, our work responds to the forceful call by Acemoğlu and Robinson (2013) to 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. While their work provides a compelling critique of the Kaldor-Hicks criterion for determining the (static) welfare consequences of trade, ours identifies the potential long run political consequences of inequality via the democratic response to macroeconomic shocks.

## 2 A Model of Protectionist Overshooting and Escalation

This section presents an overlapping generations model with endogenous dynamic political responses to external shocks. In our small-country open-economy model, two-period lived heterogeneous agents decide how much costly education to acquire during the first period of their lives, while reaping the benefits of their human capital investment in the second period. Trade policy is determined anew each period through majority voting; the decisive (median) voter at the time decides the policy for the period based on her previous human capital investment decisions and the terms of trade. Thus, the equilibrium policy outcome in each period is determined by the human capital decisions from the previous period. The central importance of the existing stock of human capital on current trade policy decisions introduces political hysteresis, even in the absence of uncertainty.[9]

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 tariff 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 the trade protection. There are then two long-run possibilities. In one scenario, rising investment in education

[8] 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.

[9] This mechanism is thus distinct from the seminal work of Fernandez and Rodrik (1991), who demonstrate the potential for political hysteresis when the identify of future winners and losers is unclear.

induces income convergence, so that as workers have time to adjust, political polarization will abate and the tariff will fall. We call this *protectionist overshooting.* Alternatively, *protectionist escalation* will occur if rising investment in education exacerbates underlying inequality. In this scenario, 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 then show that unanticipated skill-biased technological change (SBTC) is virtually isomorphic to a terms-of-trade shock in generating protectionist dynamics.

## 2.1 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, $S$. Thus, any import tariff applied on imports of the basic good $U$ 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 $\tau$ is equal to one under free trade and strictly greater than one under a tariff.[10]

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 advantages at 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 ($^y$) and the old ($^o$), comprise the total population. The population of each generation is nor-

[10] 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 $S$.

malized to one. We refer to the generation that is young at time $t$ as 'generation $t$' hereafter. Agents have rational expectations and perfect foresight.[11] Finally, we assume that tariff revenue is rebated uniformly across agents *within* each generation. This intra-generational rebating assumption intentionally removes any potential inter-generational transfer motivations for tariffs, and thus allows us to isolate the distributional motivations that are our focus.[12]

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 via costly education. Schooling takes time, and so the cost of human capital is the foregone income from work in the unskilled sector if not for time in the classroom. There are no additional pecuniary costs of education, and 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) labor 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 simple two-period overlapping generations framework the old have no incentive to acquire additional education in the second period of life. Our simple structure is thus effectively an extreme case of putty-clay skill 'stickiness' as in Matsuyama (1992).[13]

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. Producing the skill-based good requires human capital, $h$. Let an individual's output of the skill-based good, $x^s(h)$, be a linear, strictly increasing function

[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.

[12] Moreover, the absence of inter-generational transfers (together with our small-country assumption and limited suffrage) eliminates a the potential for self-fulfilling expectations equilibria. We view this as an explicit advantage of the model. See footnote 18 for further discussion.

[13] 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.

of her human capital level:

$$x^s(h) = bh(a,e) \qquad \text{with} \qquad b \geq 1, \tag{2}$$

where $b$ represents a productivity shifter that can be used to study the effect of skill-biased technological change. A type-$a$ worker's human capital in the second stage of life is strictly increasing both on her innate advantage, $a$, and the extent of education she acquired when young, $e$. We assume that education and inherent advantage are complementary in realized human capital, that is, $h(a,e)$ is super-modular in $a$ and $e$, and that the human capital return to education is strictly concave.[14] Our assumptions over human capital accumulation are summarized as follows:

**Assumption 1.**

$$\frac{\partial h(a,e)}{\partial a} > 0; \qquad \frac{\partial h(a,e)}{\partial e} > 0; \tag{3}$$

$$\frac{\partial^2 h(a,e)}{\partial a \partial e} > 0; \qquad \frac{\partial^2 h(a,e)}{\partial e^2} < 0. \tag{4}$$

**Education and Production.** Each agent chooses her education level to maximize lifetime indirect utility. Preferences are identical across individuals and functionally 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, $d_s^y(d_u^y)$ $[d_s^o(d_u^o)]$ denote the individual's consumption of good $S$ $(U)$ when she is young [old], and $u(d_u, d_s) = d_u^{(1-\alpha)} d_s^{\alpha}$. These homothetic intra-temporal preferences remove the potential for perverse income effects and allow us to focus on the skill acquisition decision abstracting from consumption smoothing,[15] With homothetic preferences, intra-period indirect utility may be written as $v(p, I) \equiv v(p)I$, where $I$ denotes current nominal income.

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

[15] Under constant marginal utility of income, agents' skill acquisition decisions are orthogonal to savings and wealth. Note that the presence of a perfect credit market would silence the effect of a consumption smoothing motive on education decisions.

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, $R_t^y$:

$$I_t^y = l_t + R_t^y = 1 - e_t + R_t^y.$$

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 production that accrues to acquired human capital, plus tariff revenue stemming from basic goods imports of the old.[16] For the young worker of generation $t$ and type $a$, income in the second period of life is given by:

$$I_{t+1}^o = 1 + bh(a, e_t)p_{t+1} + R_{t+1}^o.$$

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, the opportunity cost of education, and the future returns to human capital, every agent $a$ of each generation $t$ agent chooses her optimal level of education to solve:

$$\max_e \quad v(p_t, I_t^y) + \beta v(p_{t+1}, I_{t+1}^o), \qquad \text{or} \tag{6}$$

$$\max_e \quad v(p_t)[1 - e + R_t^y] + \beta v(p_{t+1})[1 + bh(a, e)p_{t+1} + R_{t+1}^o]. \tag{7}$$

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:

$$\beta b \frac{\partial h(a, e)}{\partial e} p_{t+1} = \frac{v(p_t)}{v(p_{t+1})}. \tag{8}$$

Using the definition of the domestic price $p_t = \frac{p_t^w}{\tau_t}$ 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 $p_t^w$ and $p_{t+1}^w$ as arguments to economize on notation.)

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

and $h_e^{-1}(\cdot)$ indicates the inverse of the first derivative of $h(a, e)$ with respect to $e$.

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

Our assumptions over human capital formation, $h(a, e)$, ensure existence and uniqueness of the optimal education function, $e(a; \tau_t, \tau_{t+1})$.[17] 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, $\beta$, and the current and (expected) future domestic relative price of the skill based good, $p_t$ and $p_{t+1}$.*

*Proof.* The signs of these effects follow directly from totally differentiating the first order condition (8), Assumption 1, and the properties of the indirect utility function:

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

where $\alpha \equiv \frac{-v_p(p_{t+1})p_{t+1}}{v(p_{t+1})}$, which by Roy's identity is the expenditure share on good $S$. □

The following corollary follows immediately from the last two inequalities, given the inverse relationship between the import tariff (applied to the basic good) and the domestic relative price of the skill-based good:

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

Intuitively, an agent's optimal education level increases with her inherent advantage due to the assumed complementarity between education and $a$. The more an agent values the future relative to the present, i.e. the greater $\beta$ is, the greater her incentive to invest in education today. Likewise, a higher domestic relative price of the skill-based good today reduces the opportunity cost of education by decreasing the real wage for unskilled work, which also increases educational attainment. Finally, a higher relative price of the skill-based good in the future increases the return to education directly.

Aggregating across all agents of both generations at a given time $t$ yields the output of each good, $q_t^s$ and $q_t^u$. (Recall that young agents provide unskilled labor only when not

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

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,* $(p_t^w, \tau_t)$ $\forall t \in \mathbb{N}$ *an* ***economic equilibrium*** *is a list of education decisions by every agent* $a \in [0,1]$*:*

$$e_t(a; \tau_t, \tau_{t+1}) = h_e^{-1}\left(a, \left(\frac{v(p_t)}{v(p_{t+1})}\frac{\tau_{t+1}}{\beta p_{t+1}^w b}\right)\right) \text{ where } p_t = \frac{p_t^w}{\tau_t} \forall t \tag{10}$$

*and associated quantities of each good produced:*

$$q_t^u = q^u(\tau_t, \tau_{t+1}) = \left(1 - \int_0^1 e_t(a; \tau_t, \tau_{t+1}) f(a) da\right) + 1 \quad \forall t \tag{11}$$

$$q_t^s = q^s(\tau_{t-1}, \tau_t) = b\int_0^1 h(a, e_{t-1}(a; \tau_{t-1}, \tau_t)) f(a) da. \quad \forall t \tag{12}$$

*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.

An *economic steady state* is then simply an economic equilibrium that obtains under a constant world price, $p^w$ and a constant tariff $\tau$ such that the domestic price $p = \frac{p^w}{\tau}$ is also constant.

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

$$e(a; \tau) = h_e^{-1}\left(a, \left(\frac{\tau}{\beta p^w b}\right)\right)\Bigg)\Bigg), \text{ where } p = \frac{p^w}{\tau}, \forall a \in [0,1] \tag{13}$$

*and constant associated output quantities*

$$q^u(\tau) = \left(1 - \int_0^1 e(a;\tau) f(a) da\right) + 1 \tag{14}$$

$$q^s(\tau) = b\int_0^1 h(a, e(a;\tau)) f(a) da \tag{15}$$

*that obtain at every period t in time.*

Finally, note that in our small open economy setting, aggregate national income is maximized under free trade; i.e. for a given $p^w$, (13) -(15) evaluated at $\tau = 1$.

## 2.2 The Political Process

We model the political process as a direct democracy over trade policy, in which only the old generation holds suffrage rights.[18] At the beginning of each period, voters choose the current period trade policy, which subsequently determines the price level and 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.

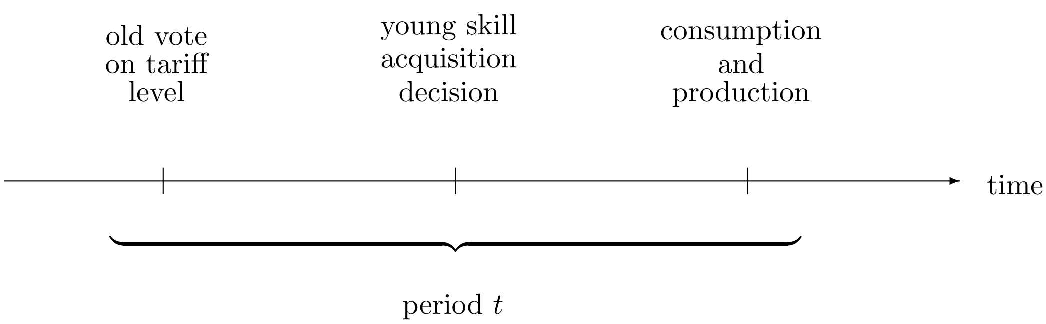



Figure 1: Within-period Sequencing.

The tariff preferences of the electorate are defined as follows. At time $t$, we denote the distribution of the (now fixed) education levels among the currently-old 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:

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

[18]By limiting suffrage 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 have no interest in the election. 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.

where $\bar{M}_t^{o,u}$ ($\bar{E}_t^{o,s}$) denotes per-capita imports of good $U$ (exports of good $S$) among the old generation at time $t$.[19] Using Roy's identity, the first order condition of the maximization problem can be written as:

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

where $E_t^{o,s}(a)$ indicates the net export position of an individual of type $a$. Rewriting again yields:

$$V_\tau^o = v_I \frac{p_t}{\tau_t} \left( \underbrace{-\Delta_t(a; \mathbf{e}_{t-1})}_{\text{individual bias}} + t_t \tau_t \underbrace{\frac{d\bar{E}_t^{o,s}}{d\tau_t}}_{(-)} \right) = 0, \tag{16}$$

where we define:

$$\Delta_t(a; \mathbf{e}_{t-1}) \equiv [E_t^{o,s}(a) - \bar{E}_t^{o,s}] = (1-\alpha)b[h(a, e_{t-1}(a)) - \int_a h(a, e_{t-1}(a)) f(a) da, \tag{17}$$

which is the net-skill position of a voter of type $a$ relative to her generation. These $\Delta$s will play a central to the remaining analysis, and so it is worth point out a couple of important properties. First, notice that $\Delta_t(a; \mathbf{e_{t-1}})$ depends on only the human capital of each voter *relative to the first moment* of the distribution of human capital in her generation, $\bar{h}_{t-1} \equiv \int_a h(a, e_{t-1}(a)) f(a) da$. Since voters are vanishingly small relative to the overall population, any one individual's education choices will not affect aggregate human capital, $\bar{h}$. Thus, we may equivalently write, $\Delta_t(a; \mathbf{e_{t-1}}) \equiv \Delta_t(a, e_{t-1}(a), \bar{h}_{t-1}(\mathbf{e_{t-1}}))$. Second, $\Delta_t(a; \mathbf{e_{t-1}})$ is fixed at the beginning of time $t$, *before* voting occurs.

The role of individual level heterogeneity is immediate from equation (16). The second term captures the typical aggregate efficiency cost of trade restrictions, which is minimized at free trade. The first term captures individual self interest. 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 ($\Delta(a; \mathbf{e}_{t-1}) > 0$) would prefer to subsidize trade. It is only a razor's edge *economically representative* 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 would vote for free trade.

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

[19] Under balanced trade, $M^{o,u} = p^w E^{o,s}$. $\bar{E}_t^{o,s} \equiv \int_a E_t^o(a) f(a) da = \int_a [x_t^{o,s}(a) - d_t^{o,s}(a)] f(a) da$ where $d_t^{o,s}(a) = (\alpha/p) I_t^o(a, \mathbf{e_{t-1}}))$ is individual $a$'s consumption of good $S$.

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 divergence between a voter's own net-skill position relative to her generation, the more she is willing to use a tariff to tilt the wage distribution in her favor 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 $t$, given implicitly by* (16)*:* $\tau_t^o(a, \mathbf{e}_{t-1}) \equiv \tau_t^o(\Delta(a, \mathbf{e}_{t-1}), \bar{E}_t^{o,s}(\mathbf{e}_{t-1}))$, *is strictly positive (negative) if and only if* $\Delta(a, \mathbf{e}_{t-1}) < 0 (> 0)$. *Moreover,* $\frac{\partial \tau_t^o(a, \mathbf{e}_{t-1})}{\partial a} < 0$, $\frac{\partial \tau_t^o(a, e_{t-1}(a), \bar{E}_t^{o,s})}{\partial e_{t-1}(a)} < 0$, $\frac{d\tau_t^o(a, \mathbf{e}_{t-1})}{da} < 0$, *and* $\frac{\partial \tau_t^o(\Delta, \bar{E}_t^{o,s})}{\partial \Delta} < 0$.

*Proof.* The first part of the proof follows directly from the first order condition in (16): evaluated at $t = 0$, $V_\tau \geq 0$ (implying a positive tariff) if and only if $\Delta \leq 0$. The second order condition, $V_{\tau\tau} < 0$ holds with strict equality (proof in Claim 1 in Appendix B). To establish the signs on the derivatives, we begin by showing that $V_{\tau a}(a, e)$ and $V_{\tau e}(a, e)$ are negative in B, Claims 2-3 in the Appendix. Then, taking the total derivative of the first order condition in (16) with respect to $a$ and $\tau$, we have that $\frac{\partial \tau_t^o}{\partial a} = -\frac{V_{\tau a}}{V_{\tau\tau}} < 0$. Likewise, the total derivative of (16) with respect to $e$ and $\tau$ yields $\frac{\partial \tau_t^o}{\partial e_{t-1}(a)} = -\frac{V_{\tau e}}{V_{\tau\tau}} < 0$. Next, $\frac{d\tau_t^o}{da} = \frac{\partial \tau_t^o}{\partial a} + \frac{\partial \tau_t^o}{\partial e(a)} \frac{\partial e(a)}{\partial a} < 0$, since the first two derivatives are negative (above) and the last factor is positive by Lemma 1. Finally, the total derivative of (16) with respect to $\Delta$ and $\tau$ 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). □

**Voting.** Trade policy is determined by majority vote. Every agent votes for her most preferred tariff policy, $\tau \in (0, \tau^P]$, where $\tau^P$ 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.[20] There is no bureaucratic or time cost of changing tariff regimes.

Political equilibrium is composed of two parts: the sequence of tariffs as a function of education, and the sequence of education decisions as a function of tariffs. As before, equilibrium education is determined by current and expected prices under rational expectations according to (9). 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

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

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 time $t$ is independent of future trade policy.[21] 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 $t$.[22]

Using $\Delta_t^m \equiv \Delta_t(a^m; \mathbf{e_{t-1}})$ to denote the relative net skill 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 $(\tau_t, \Delta_t^m, \mathbf{e_{t-1}})$ triples such that starting from $\mathbf{e_0}$ the following holds for all $t \in \mathbb{N} = \{1, ...\infty\}$:*

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

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

3. $\Delta_t^m = \Delta(a^m, \mathbf{e}_{t-1}) = (1-\alpha)b[h(a^m, e(a^m; \tau_{t-1}, \tau_t)) - \int_a h(a, e(a; \tau_{t-1}, \tau_t))f(a)da]$,

*where $p_t = \frac{p_t^w}{\tau_t} \forall t$ and $V^o(\cdot)$ denotes the indirect utility for a voter in the second period of her life at time $t$.*

The first condition requires that the equilibrium realized tariff maximizes the indirect utility of the (older) median voter in every period, $t$. The second condition requires that individuals' skill acquisition strategies are optimal under rational expectations of the current and future equilibrium tariffs.

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.

[21] 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.

[22] 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 $\int_a h(a, e_{t-1}(a))f(a)da$, which enters both $\bar{E}_{t-1}^{o,s}$ and enters $\Delta_{t-1} = (1-\alpha)b[h(a, e_{t-1}(a)) - \int_a h(a, e_{t-1}(a))f(a)da]$.

**Definition 4.** *A **political steady state**, summarized by* $(\tilde{\tau}, \tilde{\Delta}^m, \tilde{\mathbf{e}})$ *is characterized by equations* (13) – (15) *and a sequence of constant tariffs* $(\tau_t = \tau)_{t\in\mathbb{N}}$ *that jointly satisfy Definition 3:*

$$\begin{aligned} \tilde{\tau} &= \arg\max_{\tau} \quad V^o(a^m, \tilde{\mathbf{e}}; \tau), && (18) \\ e(a, \tilde{\tau}) &= h_e^{-1}\left(a, \left(\frac{\tilde{\tau}}{\beta p^w b}\right)\right)\forall a && (19) \\ \tilde{\Delta}^m &\equiv \Delta(a^m, \mathbf{e}(\tilde{\tau})), && (20) \end{aligned}$$

**Steady State Properties.** An unique interior steady state exists if there is one (and only one) fixed point solution to equations (18)-(20) such that $\tilde{\tau} \leq \tau^P$ (i.e., a non-prohibitive tariff). For the remainder of this paper, we focus on the case in which the steady state equilibrium relative net-skill position of the median voter is negative: $\tilde{\Delta}^m < 0$, which is guaranteed if even under free trade the distribution of the returns to human capital is skewed toward the top.

We assume the following (sufficient) conditions for a 'nice' case in which there is an unique, stable, interior political steady state. Intuitively, uniqueness and stability require that the steady state tariff is not overly responsive to small changes in $\tilde{\Delta}^m$. As we explore below, the steady state position of the median voter, $\tilde{\Delta}^m$ may be increasing or decreasing in the tariff level; this stability condition permits either case (and so does not rule out economically interesting cases).

**Assumption 2.** *Unique, Stable, Interior Steady State:*

$$\begin{gathered} \left(\frac{h_e^2}{h_{ee}}\Big|_{a^m} - \int_a \frac{h_e^2}{h_{ee}} f(a)da\right) < \frac{\tau^2}{p^w}\left(\bar{h} + \frac{1}{bp^w}\right)(\alpha(\tau - \alpha t)), \\ \Delta^m(\tau)\Big|_{\tau=1} < 0 \textit{ and } \Delta^m(\tau)\Big|_{\tau^P} > \tau^{-1}(\Delta^m)\Big|_{\tau^P}. \end{gathered} \tag{21}$$

As we show in the appendix, Assumption 2 ensures uniqueness and stability for not only the steady state equilibrium, but also for every out of steady state fixed point equilibrium.

Finally, note that hereafter we economize on notation by referring just the equilibrium and steady state *pair*, $(\Delta^m, \tau)$, to stand in for the triple $(\Delta^m, \tau, \mathbf{e})$, taking for granted that the optimal education decisions are implied by the underlying model (e.g. condition 2 in Definition 3).

**Lemma 3.** *Under Assumption 2, the period* $t$ *equilibrium pair* $(\Delta_t^m, \tau_t)\forall t$ *is unique and stable, both in and out of steady state.*

*Proof.* The first condition in Assumption 2 ensures that the tariff is not decreasing more quickly in $\Delta$ than $\Delta$ is falling in $\tau$: $\frac{d\Delta^m(\tau)}{d\tau} > \left.\frac{d\Delta^m}{d\tau^o}\right|_{\tau^o=\tau(\Delta^m)}$, which implies uniqueness and stability. The second set of conditions imply interiority. See Appendix C for formal treatment. □

To build intuition, Figure 2 depicts steady state where the median voter's education level depends on tariffs, and tariffs depend on the median voter's level of education (relative to the average in the economy). Both functions are (unambigously) downward sloping in $(e^m, \tau)$ space: the median voter's steady state education schedule is decreasing in the tariff (Lemma 1), and the steady state equilibrium tariff is decreasing in the median voter's steady state education level (keeping the mean level of human capital fixed) (Lemma 2). The steady state equilibrium median voter education, trade policy pair is labeled $(\tilde{e}^m, \tilde{\tau})$. Notice that the tariff locus passes through a free trade a benchmark education level $\hat{e}$, at which the median voter would prefer free trade (as a function of the overall education level of the population).

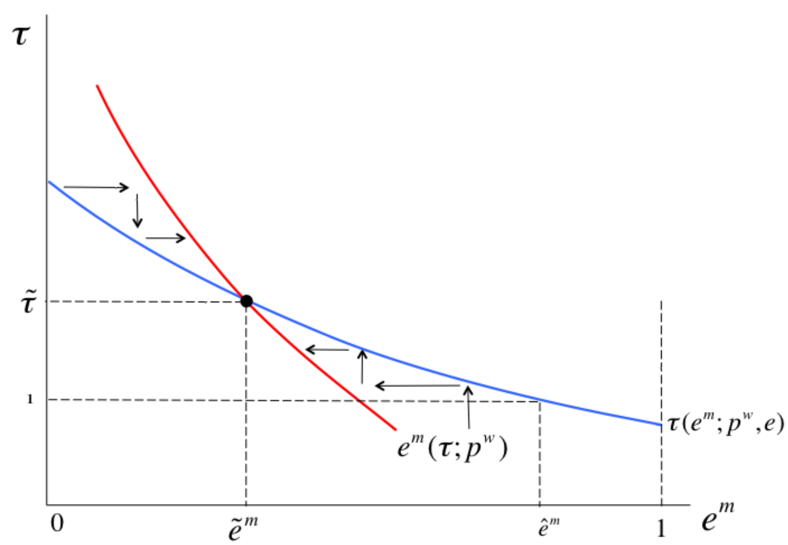



Figure 2: Steady State

Alternatively, one can depict equilibrium in $(\Delta^m, \tau)$ space, as we do in the appendix. We present graphs in terms of $e^m$ and $\tau$ simply to highlight the intuition.

# 3 Policy Response to Exogenous Shocks

We now examine the short and long run consequences of a sharp, unexpected, permanent increase in the world relative price of the skilled good. We adopt the perspective of a relatively skill-abundant country in which the initial steady state distribution of human capital is assumed initially to be skewed toward the top (i.e. $\tilde{\Delta}^m < 0$). This thought experiment 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 country like the United States. In an extension, we show that a skill-augmenting technology shock is virtually isomorphic in its political consequences.

## 3.1 An Unanticipated Increase in the Terms-of-Trade, $p^w$

Starting from a political steady state summarized by $(\tilde{\Delta}^m, \tilde{\tau}; p^{wo})$ where $\tilde{\Delta}^m < 0$, consider an unanticipated jump in the world price to $p^{w\prime} > p^{wo}$ at time $t = T$.[23] The increase in the relative world price of the skilled good will change both the incentives to acquire education and also 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 continue to assume the necessary regularity conditions to ensure an unique, stable steady state.

**Steady State.** We begin by examining the effect of the terms-of-trade shock on the steady state tariff function, holding the underlying distribution of education (and therefore $\Delta^m$) fixed.[24] Below, we show that an increase in the terms of trade, $p^w$, will further polarize tariff preferences, but not by so much that voters would fully offset the increase in $p^w$. Thus, the net effect of an increase in $p^w$ will be at least a small increase in the domestic price, $p$. Formally:

**Lemma 4.** *Polarization effect of an increase in $p^w$. For any $p^{w\prime} > p^{wo}$ and any $\Delta^m < 0$:*

1. $\tau(\Delta^m; p^{w\prime}) > \tau(\Delta^m; p^{wo})$ *iff* $\Delta^m < 0$.
2. $\tau(\Delta^m; p^{w\prime}) < \tau^{FC}$ *where* $\tau^{FC} \equiv \frac{p^{w\prime}}{p^{wo}} \tau(\Delta^m; p^{wo})$.

[23]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). Since this would not change our results, we have chosen to use this simpler set up.

[24]Recall that $\Delta^m \equiv \Delta(a^m, \mathbf{e}) = (1-\alpha)b[h(a, e(a)) - \bar{h}(a, \mathbf{e})]$.

*Proof.* Part (*i*): Totally differentiating the first order condition of the optimal tariff function in (16) with respect to $\tau$ and $p^w$ yields $\frac{d\tau^o}{dp^w} = -\frac{V_{\tau p^w}}{V_{\tau\tau}}$. 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 $\Delta^m < 0$ ($\geq 0$), which yields the result. For part (*ii*), it is sufficient to show that evaluated at the fully compensating tariff, $V_\tau(\Delta^m; p^{w\prime})$ is strictly less than zero (which implies that the median voter would prefer a strictly smaller tariff). Evaluating the first order condition of the optimal tariff problem at the new terms of trade and $\tau^{FC}$, we have $V_\tau(\Delta^m; p^{w\prime})\Big|_{\tau^{FC}} = v_I \frac{p'}{\tau^{FC}}\big( - \Delta^m + t^{FC}\tau^{FC}\frac{dE^{FC}}{d\tau}\big)$. For any given $\Delta^m$, the initially optimal tariff, $\tau^o(\Delta^m)$ is given by the first order condition $\Delta^m = t^o\tau^o\frac{dE^o}{d\tau}$. Substituting in, and using that $\tau^{FC}$ holds the domestic price fixed at the initial level by definition ($p' = p^o$), we have: $V_\tau(\Delta^m; p^{w\prime}) = v_I\frac{p^o}{\tau^{FC}}\big( - t^o\tau^o\frac{dE^o}{d\tau} + t^{FC}\tau^{FC}\frac{dE^{FC}}{d\tau}\big)$. 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: $V_\tau(\Delta^m; p^{w\prime})\Big|_{\tau^{FC}} < 0 \Rightarrow \tau(\Delta^m; p^{w\prime}) < \tau^{FC}$. $\square$

This polarization result makes sense. Voters choose trade policy to balance their individual incentive to tilt the domestic relative price in their favor against the distortionary cost of trade restrictions. In the initial steady state, these two forces are exactly equal for the median voter. When the world price increases, this balance is disrupted. Holding the current tariff fixed, an increase in the world relative price would (strictly) increase the redistributive motive to change the tariff relative to the distortionary cost. Thus, a relatively less skilled, $\Delta^m < 0$ median voter will increase the tariff at least a little bit. (The opposite would of course be true if $\Delta^m > 0$.) The same logic can be used to show that this increase in the tariff will less than fully offset the terms-of-trade change. If the median voter were to hold the domestic price fixed by implementing a fully compensating tariff, the distortionary cost of the tariff would strictly increase while the redistributive motive would stay the same.

This lemma then allows us to put bounds on the new steady state possibilities as follows:

**Proposition 2.** *Steady State response to an increase in $p^w$. Compared to an initial steady state summarized by $(\tilde{\Delta}^m, \tilde{\tau}; p^{wo})$ where $\tilde{\Delta}^m < 0$ and $\tilde{\tau} < \tau^P$, the new steady state under a higher world price $p^{w\prime} > p^{wo}$, $(\tilde{\tilde{\Delta}}^m, \tilde{\tilde{\tau}}; p^{w\prime})$ has the following properties:*

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

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

*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 $\tilde{\Delta}^m$, is strictly less than fully compensating; and $(b)$ the value of the new steady state $\Delta^m$ function, evaluated at the fully compensating tariff, coincides with the original steady state $\tilde{\Delta}^m$. 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 $\tau$, holding $p$ fixed. Since by definition $\tau^{FC}$ would hold the domestic price unchanged, $\Delta^m(\tau^{FC}; p^{w\prime}) = \Delta^m(\tilde{p}) = \tilde{\Delta}^m$. 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 ($\tilde{\tilde{p}} > \tilde{p}$). 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. □

Figure 3 illustrates. Starting from an initial steady state at $(\tilde{e}^m, \tilde{\tau})$, an increase in $p^w$ will cause the steady state education locus $e^m(\tau)$ to shift rightward for all values of $\tau$: 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 $\tau(e^m; \mathbf{e})$ 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 2 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 $\tilde{\tilde{e}}^m \geq \tilde{e}^m$, the new steady state tariff $\tilde{\tilde{\tau}} < \tau^{FC}$ may be above or below the initial tariff, $\tilde{\tau}$ depending on where new steady state loci intersect.

**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 (with an abuse of notation), $\Delta^m_T = \tilde{\Delta}^m = \Delta(\tilde{e}^m, \tilde{\mathbf{e}})$. This

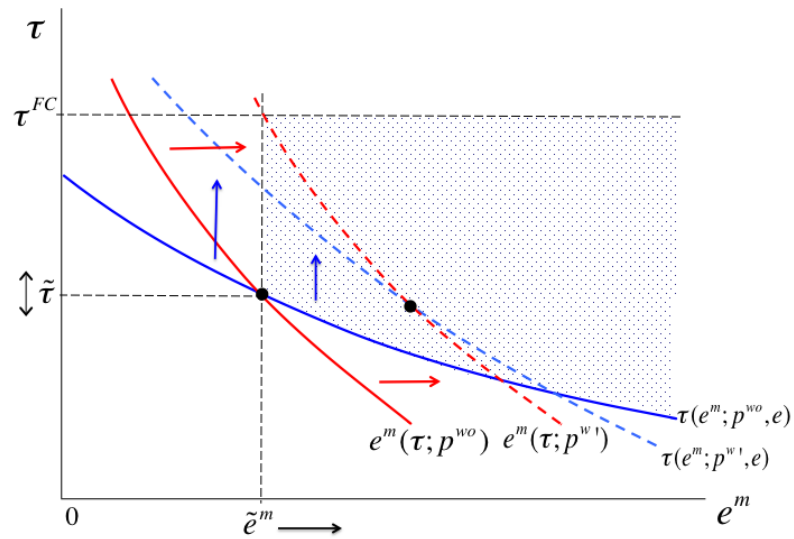



Figure 3: Steady State Response to $p^w \uparrow$

serves as the starting value of 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 vote on trade policy. Notice that the optimal tariff function $\tau(\Delta^m)$ depends only on the concurrent $\Delta^m$, and is therefore the same at the time of the shock as in the new steady state: $\tau_T = \tau(\tilde{\Delta}^m, p^{w\prime})$.[25] Given our initial assumption that the returns to human capital are skewed toward the top ($\tilde{\Delta}^m < 0$), 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 3.** *Protectionist Surge: Starting from an initial steady state summarized by* $\{\tilde{\Delta}^m, \tilde{\tau}; p^{wo}\}$ *where* $\tilde{\Delta}^m < 0$ *and* $\tilde{\tau} < \tau^P$*, 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.* $\tau_T > \tilde{\tau}$ *and* $p_T > \tilde{p}$*, where* $\tilde{p} \equiv \frac{p^{wo}}{\tilde{\tau}}$.

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

Given the increase in the domestic relative price of the skill-intensive good at time $T$,

[25]The same is not true, however, of the out-of-steady-state tariff locus in $(e^m_{t-1}, \tau_t)$ space, since $\tau_t = \tau(e^m_{t-1}, \mathbf{e_{t-1}}; p^w)$. Thus, the tariff locus (as a function of $e^m_{t-1}$) will shift over time with the aggregate lagged education level.

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 $\tau_{t+1}$. 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 is pinned down by $\Delta_T^m$ and the second is endogenous: $\tau_{T+1} = \tau(\Delta_{T+1}^m; p^{w\prime}) = \tau(e_T^m, \mathbf{e_T}; p^{w\prime})$.

Under rational expectations, the expected future tariff will 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.[26] Our regularity assumptions assure a unique fixed point solution in each period, so that transition is pinned down by parameters. But depending on the underlying functional form assumptions, there are two possibilities for how this transition will evolve.

Intuitively, if young voters expect tariff liberalization 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 – that is, if rising education causes $\Delta^m$ to converge toward zero – so that the median will in fact be less protectionist in the future. This may, but 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 $\Delta_{T+1}^m < \Delta_T^m$, then $\tau_{T+1} > \tau_T$: the median voter will even more protectionist following the shock.

We call this first possibility *Protectionist Overshooting*: following an initial tariff surge at the time of the shock, trade policy will gradually liberalize as workers acquire more education, the distribution of human capital converges, 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)

[26]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, all agents understand that the median individual, $a^m$, will necessarily be the median voter with respect to trade policy in the subsequent period.

Below, we show that whether the protectionist surge will be followed by tariff escalation or a gradual liberalization depends on whether an increase in the skill premium causes divergence or convergence in the endogenous distribution of human capital: i.e. $\frac{d\Delta^m(p)}{dp} \gtrless 0$.

**Proposition 4.** *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* $\tilde{\tau}$ *to* $\tau_T$*;*

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

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

*Proof.* Proposition 3, establishes point $(i)$ directly. We also use it to prove part $(ii)$. At the time of the shock, $\Delta^m_T = \tilde{\Delta^m}$ and $\tau_T = \tau(\tilde{\Delta^m}; p^{w\prime}) > \tau(\tilde{\Delta^m}; p^{wo})$. As established as part of the proof for Proposition 3, the new steady state $\Delta^m(\tau; p^{w\prime})$ locus takes a value of $\tilde{\Delta^m}$ 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 $\tau$. Since $\tau(\Delta^m; p^{w\prime})$ is also (always) decreasing in $\Delta^m$, the new steady state $\Delta^m(\tau)$ and $\tau(\Delta^m)$ schedules must intersect at some value where $\Delta^m > \tilde{\Delta}^m = \Delta^m_T$ and $\tau(\tilde{\tilde{\Delta}}^m; p^{w\prime}) < \tau_T$.

To establish part $(iii)$ of the proposition, we use induction to trace out the fixed point equilibrium values of $\tau$ and $\Delta^m$ 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 $\Delta^m(p_T) > \Delta^m_T = \tilde{\Delta^m} = \Delta^m(\tilde{p})$, since $p_T > \tilde{p}$ (by Proposition 3) and $\Delta^{m\prime}(p) > 0$ by assumption. This candidate cannot be a steady state, however, since then we would have $\tau_{T+1} = \tau(\Delta^m(p_T); p^{w\prime}) < \tau_T = \tau(\Delta^m(\tilde{p}); p^{w\prime})$, resulting in a contradiction. Now let $\Delta^m_{T+1} = \Delta^m(p^{w\prime}/\tau_T, p^{w\prime}/\tau_{T+1})$, where $\tau_{T+1} = \tau(\Delta^m_{T+1}; p^{w\prime})$. Compared to our benchmark, it must be that $\tau_{T+1} < \tau_T$ and $\Delta^m_{T+1} > \Delta^m(p^{w\prime}/\tau_T, p^{w\prime}/\tau_T) = \Delta^m(p_T) > \Delta^m_T$, since, according to our regularity conditions, $\Delta^m(\tau_{t-1}, \tau_t)$ decreases faster in its second argument than $\tau(\Delta^m_t; p^{w\prime})$ decreases in $\Delta^m$. 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. □

Figure 4 offers a simple illustration in $(e^m, \tau)$ space. At the time of the shock, the tariff schedule pivots around the initial value of $\hat{e}(\tilde{\mathbf{e}})$, leading to an immediate and unambiguous increase in the tariff to $\tau_T$, while the education of the (then-old) median voter remains fixed at $\tilde{e}^m$. (With a slight abuse of notation, we label this out-of-steady-state tariff locus $\tau^T(\tilde{e}^m)$.) Over time, the out-of-steady state education and tariff loci (not shown) will gradually shift rightward as the domestic price and overall (mean) education level rise, following a monotonic decline in tariffs and increase in $e^m$ to the new steady state. We should emphasize that this graphical depiction in $(e^m, \tau)$ space is meant to serve as illustration only. While $e^m$ is implicit, the formal proof (above) uses $\tau_t = \tau(\Delta_t^m)$ and $\Delta_t^m(\tau_{t-1}, \tau_t)$. In Appendix A, we offer the more technical graphical exposition in $(\Delta_t^m, \tau_t)$ space for the interested reader.

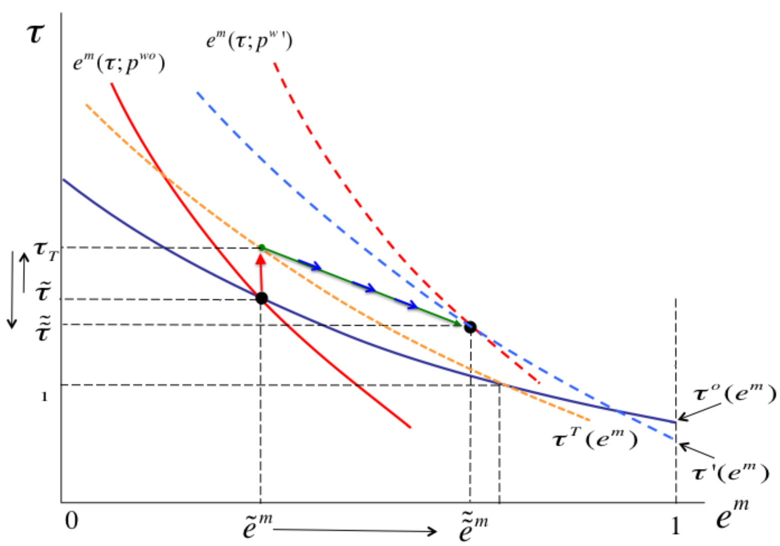



Figure 4: Protectionist Overshooting

Figure 5 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 – but either way, 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. At the same time, the leftmost panel highlights a particularly stark prediction. Even if a terms-of-trade shock will ultimately result in lower tariffs, the short run response points in exactly the opposite direction: a "rosy" long run is preceded by a rocky transition.

Viewing the transition dynamics through the lens domestic prices reveals that the protectionist surge at time $T$ is acting as a shock absorber for the overall economy. The

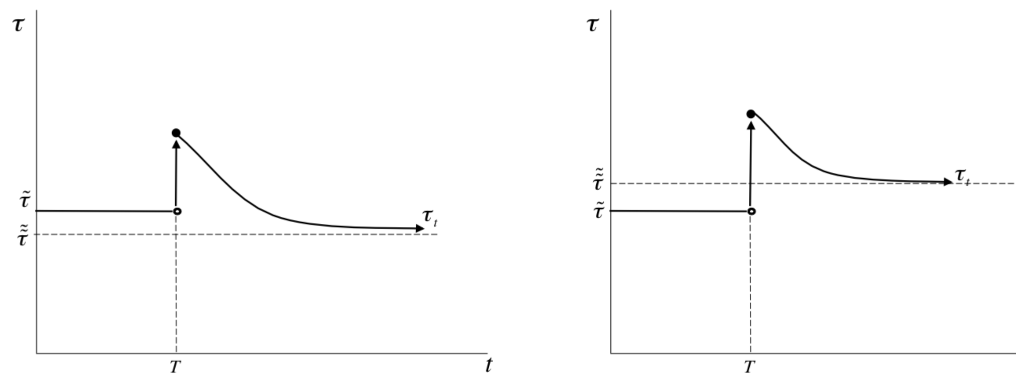



Figure 5: Protectionist Overshooting: Dynamic Tariff Response to $p^w \uparrow$ if $\Delta^{m\prime}(p) > 0$.

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. Figure 6 illustrates.

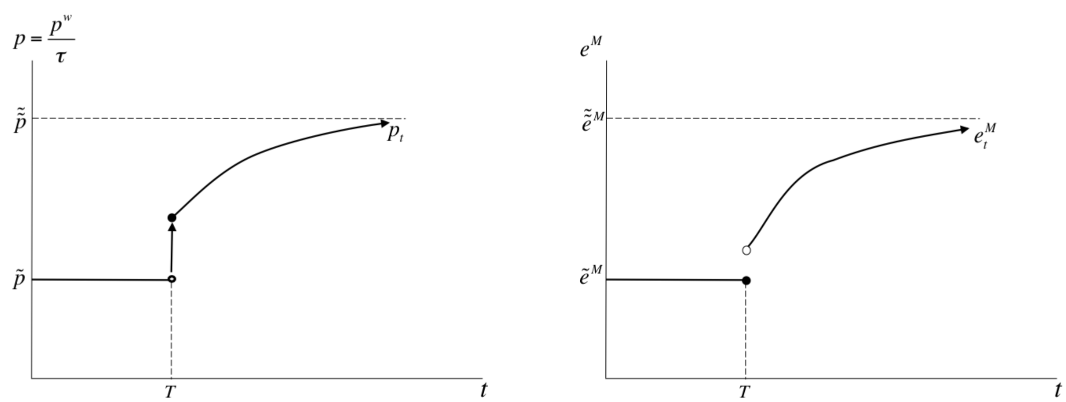



Figure 6: Overshooting Case: Time Path for Prices and Education

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 $T$, using future gains to compensate the old generation at time $T$. Unfortunately, fully-compensating, time-consistent intergenerational transfer schemes are notoriously difficult to engineer and implement. Absent such a possibility, welfare-improving policy interventions would speed the pace of adjustment by reducing labor market frictions, or reduce the underlying economic inequality that is responsible for both the protectionist surge and the deviation in steady state tariffs from free trade. We explore these possibilities further in Section 4.

We now turn to examine the alternative case, in which a rising skill-premium exacerbates underlying inequality.

**Proposition 5.** *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* $\tilde{\tau}$ *to* $\tau_T$*;*

*ii) a new steady state characterized by* $\tilde{\tilde{\tau}} \geq \tau_T > \tilde{\tau}$ *and* $\tilde{\tilde{\Delta}}^m < \tilde{\Delta}^m$*; 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.* As in the previous proof, point $(i)$ is established directly by Proposition 3, which we also use to prove part $(ii)$. Here again, we use that the new steady state locus $\Delta^m(\tau; p^{w\prime})$ takes the value of $\tilde{\Delta}^m$ evaluated at $\tau_{FC}$, whereas the new steady state tariff locus evaluated at $\tilde{\Delta}^m$ takes a strictly smaller value: i.e. $\tau(\tilde{\Delta}^m; p^{w\prime}) < \tau^{FC}$. Under the assumption that $\frac{d\Delta^m(p)}{dp} < 0$, the steady state $\Delta^m(\tau; p^{w\prime})$ schedule is *increasing* in $\tau$. Since $\tau(\Delta^m; p^{w\prime})$ is always decreasing in $\Delta^m$, this implies that the steady state $\Delta^m(\tau)$ and $\tau(\Delta^m)$ schedules must intersect at some value $\tilde{\tilde{\Delta}}^m < \tilde{\Delta}^m$ and $\tilde{\tilde{\tau}} > \tau_T > \tilde{\tau}$. In the razor's edge case in which $\frac{d\Delta^m(p)}{dp} = 0$, transition will be instantaneous at time $T$ and $\tilde{\tilde{\Delta}}^m = \tilde{\Delta}^m$ and $\tilde{\tilde{\tau}} = \tau_T > \tilde{\tau}$.

To establish part $(iii)$ of the proposition, $\Delta^m$ and $\tau$ (and hence $p$) oscillate around the new steady state, and converge to 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 $\Delta^m$ is decreasing in $p$ by assumption, this would imply $\Delta^m_{T+1} \equiv \Delta^m(p_T, p_{T+1}) < \Delta^m(p_T, p_T) \equiv \Delta^m(p_T) < \tilde{\Delta}^m$. But since the tariff schedule $\tau(\Delta^m; p^w)$ is decreasing in $\Delta^m$, this would then imply that $\tau(\Delta^m_{T+1}; p^{w\prime}) > \tau(\tilde{\Delta}^m; p^{w\prime}) = \tau_T$, which is a contradiction. It must also be true that $\tau_{T+1} \geq \tilde{\tilde{\tau}}$. Again, suppose not: i.e. let $\tau_{T+1} < \tilde{\tilde{\tau}}$, which would imply that $p_{T+1} > \tilde{\tilde{p}}$. Since $p_T > \tilde{\tilde{p}}$ from part $ii)$ of the proposition, it would then be the case that $\Delta^m(p_T, p_{T+1}) < \tilde{\tilde{\Delta}}^m$, which would in turn imply that $\tau_{T+1} > \tilde{\tilde{\tau}}$: another contradiction. Thus, $\tau_{T+1} \geq \tilde{\tilde{\tau}} \geq \tau_T$ and $\Delta^m_{T+1} > \tilde{\tilde{\Delta}}^m > \tilde{\Delta}^m$

We can follow the same procedure to show that $\tau_{T+2} \leq \tilde{\tilde{\tau}} \leq \tau_{T+1}$. Suppose not. This would then imply that both $\tau_{T+1}, \tau_{T+2} > \tilde{\tilde{\tau}}$, so that $p_{T+2}, p_{T+1} < \tilde{\tilde{p}}$. But then, $\Delta^m_{T+2} \equiv \Delta^m(p_{T+1}, p_{T+2}) < \tilde{\tilde{\Delta}}^m$. And since the tariff is decreasing in $\Delta^m$, this would mean that $\tau(\Delta^m_{T+2}; p^{w\prime}) > \tilde{\tilde{\tau}}$: contradiction. Thus, it must be true that $\tau_{T+2} \leq \tilde{\tilde{\tau}} \leq \tau_{T+1}$ and likewise $\Delta^m_{T+2} \leq \tilde{\tilde{\tau}} \leq \tau_{T+1}$ 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 $(\Delta^m_T, \tau_T)$. Next consider $(\Delta^m(p_T, p_T), \tau(\Delta^m(p_T, p_T))$ which lies on the new $\tau$-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 $\Delta^m$-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 $\tau$-schedule is less than the slope of the new steady state $\Delta^m$-schedule, it must hold that $(\Delta^m_{T+2}, \tau_{T+2})$ is closer to the new steady state than $(\Delta^m_T, \tau_T)$. We can repeat this argument for all successive full oscillations, which establishes the result.

□

As before, the interested reader is referred to Appendix A for a detailed graphical exposition. Figure 7 offers a simpler illustration in $(e^m, \tau)$ 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 loci (not shown), $e(\tau_{t-1}, \tau_t)$, shift toward the new steady locus, following an oscillating convergence pattern. The out of steady state tariff locus behaves similarly. Over time, both loci converge to the new steady state as the tariff swings gradually dissipate.

The case of protectionist escalation highlights an important political tension that arises when education and inequality move together. When an increase in domestic skill premium induces educational investments that *increase* economic inequality (i.e. $\Delta^{m\prime}(p) < 0$), the (unambiguous) increase in $p_T$ at the time of the shock (Proposition 3) boosts human capital more at the top of the income distribution than at the bottom. The result is growing inequality coupled with the rise in overall education, which will make the next median voter more protectionist than her predecessor. The political pendulum will then swing the other way: when the median voter at time $T+1$ raises the tariff, she will drive down inequality (and education), leaving her successor in period $T+2$ somewhat less protectionist. (And so on.) Gradually, the swings in tariffs and human capital will moderate, converging to a new steady state.

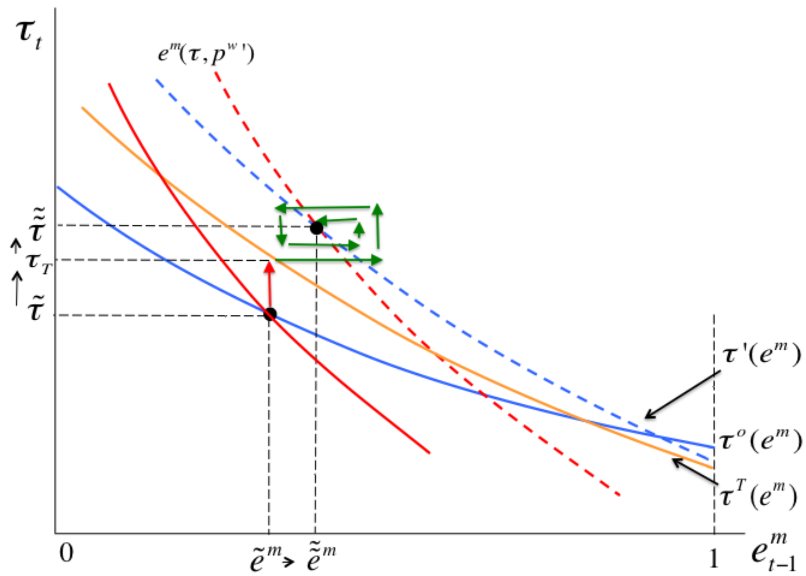



Figure 7: Protectionist Escalation

### 3.2 Technology Shocks

Here we show that the political implications of (unanticipated) skill-biased technological change (SBTC) can mimic the effects of a terms-of-trade shock. Formally, we consider the effect of a permanent, unanticipated increase in the relative productivity of skilled labor, which is summarized by the parameter $b$ in our baseline model. While the underlying mechanics are different, the political effects are essentially the same.

**Proposition 6.** *Polarizing effect of SBTC. Starting from an initial steady state summarized by $\{\tilde{\Delta}^m, \tilde{\tau}; p^{wo}\}$ where $\tilde{\Delta}^m < 0$ and $\tilde{\tau} < \tau^P$, an unanticipated skill-augmenting technological improvement that increases $b^o$ to $b' > b^o$ at time $t = T$ leads to:*

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

*ii) followed subsequently by either:*

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

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

*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: $\Delta_T^m = (1-\alpha)b'[\tilde{h}(a^m) - \int_a \tilde{h}(a)f(a)da] = \frac{b'}{b^o}\tilde{\Delta}^m$. Since $\tilde{\Delta}^m < 0$ it must be that $\Delta_T^m < \tilde{\Delta}^m$. 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, $\frac{d\tau}{db}\Big|_{\tau^o} = -\frac{V_{\tau b}}{V_{\tau\tau}}$. Recall from Claim 1 that $V_{\tau\tau}\Big|_{\tau^o} < 0$. Thus, the sign of $\frac{d\tau}{db}$ is give by the sign of $V_{\tau b}(\tau, b)\Big|_{\tau^o}$. Claim 6 in the Appendix proves that, holding education fixed, $V_{\tau b}\Big|_{\tau^o} > 0 \iff \tilde{\Delta}^m < 0$, which establishes part ($i$) of the proposition.

Subsequent to time $T$, there are two possibilities depending on the sign of $\Delta^{m\prime}(p)$. First, consider the "overshooting" case ($a$), in which $\Delta^{m\prime}(p) > 0$. We need only to show that $\Delta_{T+1}^m > \Delta_T^m$, after which the transition proceeds via monotonic tariff adjustment just as in Proposition 4. Using the definition of $\Delta^m(\cdot)$ we first show that $\Delta^m(\tau_T, \tau_T; b') > \Delta_T^m$: $\Delta^m(\tau_T, \tau_T; b') = \Delta_T^m + (1-\alpha)b'\Big[\big(\frac{dh(a^m)}{db}\big) - \big(\frac{d\bar{h}}{db}\big)\Big] = \Delta_T^m + (1-\alpha)\Big[\big(\frac{h_e^2}{-h_{ee}}\big)\Big|_{a^m} - \int_a \big(\frac{h_e^2}{-h_{ee}}\big)f(a)da\Big]$. But $\Delta^{m\prime}(p) >$ implies that $\big(\frac{h_e^2}{-h_{ee}}\big)\Big|_{a^m} > \int_a \big(\frac{h_e^2}{-h_{ee}}\big)f(a)da$. Thus, it must hold that $\Delta^m(\tau_T, \tau_T; b') > \Delta_T^m$. Then, since $\tau_t(\Delta_t^m)$ is also decreasing in $\Delta^m$, the fixed point intersection of $\tau(\Delta_{T+1}^m)$ and $\Delta^m(\tau_T, \tau_{T+1})$ must occur for some $\Delta_{T+1}^m > \Delta_T^m$ and $\tau_{T+1} < \tau_T$ .

We use a similar technique for case ($b$). Again, we need to show only that $\Delta_{T+1}^m \leq \Delta_T^m$, after which transition will proceed via the same oscillating tariff pattern in Proposition 5. (As before, under the razor's edge case in which $\Delta^{m\prime}(p) = 0$, transition will be immediate.) Applying the same logic as above, we have $\Delta^m(\tau_T, \tau_T; b') = \Delta_T^m + (1-\alpha)\Big[\big(\frac{h_e^2}{-h_{ee}}\big)\Big|_{a^m} - \int_a \big(\frac{h_e^2}{-h_{ee}}\big)f(a)da\Big]$. But when $\Delta^{m\prime}(p) \leq 0$ the second term is (weakly) negative, so that $\Delta^m(\tau_T, \tau_T; b') < \Delta_T^m$. Since $\Delta^m(\tau_T, \tau_{T+1})$ is now increasing in the second argument, it must hold that $\Delta_{T+1}^m \leq \Delta_T^m$ and $\tau_{T+1} \geq \tau_T$. □

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;[27] Proposition 6 implies that the root cause may be immaterial politically. Whether caused by technology or trade, the democratic consequences of rising inequality for tariffs could be the same.

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

# 4 Discussion

In this section, we use the model to discuss how domestic policies and multilateral trade rules may defuse or exacerbate protectionist pressure in the long and short run.[28] 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.

First, popular support for protectionism falls when individual voters' incentives are more closely aligned with the overall economy. Recall that the first order condition in (16) implicitly defines the most preferred tariff for a voter as a function of his relative net skill position:

$$t_t \tau_t = \frac{\Delta_t(a)}{d\bar{E}_t / d\tau_t} \geq 0 \iff \Delta_t(a) \leq 0 \tag{22}$$

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 the individual bias $|\Delta(a)|$ for a sufficiently large set of politically decisive voters.[29] As long as some part of individuals' earnings are derived from market wages, and as long as there is underlying inequality in 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 "basic income" – may have a limited direct effect 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 suggests that the transfer scheme would have *no* effect on the optimal tariff, since the transfers would not enter the first order condition in (22) at all.[30] If instead transfers depended on prices but not individual characteristics (for instance, via national income), they would reduce but not eliminate the influence of individual bias in tariff preferences (by adding an additional term to the denominator in (22)). A critical caveat: to the extent that unconditional redistribu-

[28] 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.

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

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

tion increases workers' labor market flexibility, such a scheme may reduce the dispersion of $\Delta(a)$, which would reduce both inequality and protectionist pressure. Pathbreaking 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. More generally, to completely eliminate individual bias from influencing tariffs, any tax and transfer scheme would need to depend on both domestic prices and $a$ to exactly offset $\Delta(a)$.

This leads to our second point: *conditional* redistribution policies that successfully reduce the individual bias 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. 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 (16), but not the first order condition governing their optimal educational decisions in (9). This 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 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.[31]

The third implication of our model for domestic policy is perhaps the most important: 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 $\Delta'(p)$. If a falling relative wage for low-skilled work induces individuals at the bottom of the income

[31] See Blanchard and Willmann (2016) for more on the interaction between educational reform and trade.

distribution to increase education faster than the mean so that $\Delta'(p) > 0$, 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 $\Delta'(p) < 0$, 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 under the initial steady state. Domestic economic policies can influence the sign of $\Delta'(p)$. 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 would have the opposite effect and could even shift the long run equilibrium from protectionist overshooting to escalation.

Fourth, we note the simple but important point that reducing labor market frictions among voters at the time of the shock will speed transition to the new steady state. Allowing the young to vote would be one such mechanism.[32] 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 and more efficient. More generally, any reduction in the frictions that limit workers' ability to respond to a changing national labor market should weakly increase support for globalization.[33]

Finally, turning from domestic policy implications to multilateral, our model offers an argument in support of including escape clauses (safeguards) in trade agreements. Here we focus on the case of protectionist overshooting depicted in Figure 4, 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 $\tilde{\tilde{\tau}} < \tilde{\tau}$. Thus, if allowed to run its course, the 'home' country's tariff would eventually fall below the original stead state level, leaving both this country and (if it is large in world markets) its trading partners better off.[34]

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

[32]Outside the model, the broader implication is then that increasing voter turnout among the younger generations would speed transition.

[33]This point is essentially an application of Le Chatelier's principle, per Varian (1992).

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

trade war.[35] 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 $\tilde{\tilde{\tau}}$ would be lost.

## 5 Empirical Context

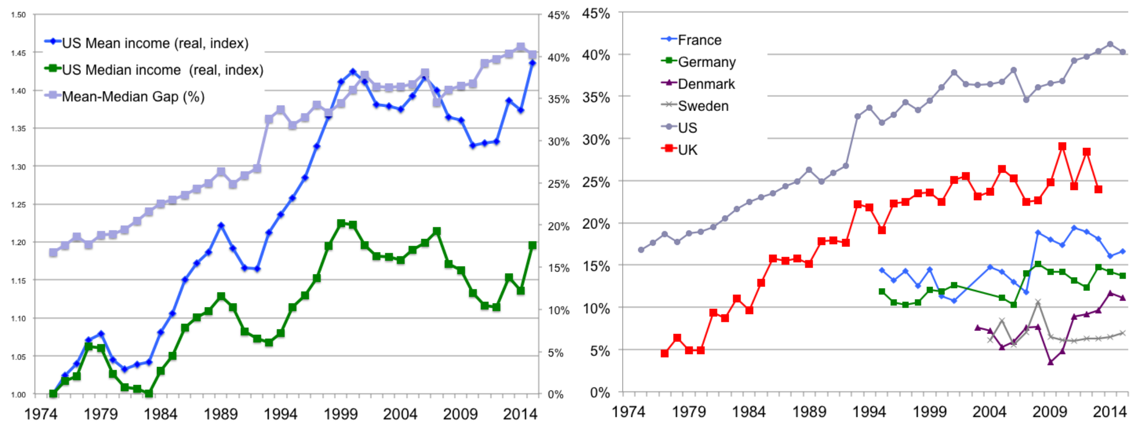
Our theoretical exercise highlights two basic insights that can be used to assess the potential for protectionist surges and their long-run consequences. First, greater dispersion in individuals' returns to economic openness will result in larger and longer-lasting protectionist reactions to a terms-of-trade 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 long term welfare.

Below, we present anecdotal evidence on these two drivers of protectionism. Our model makes a crucial distinction between education and inequality: how a country responds to macroeconomic shocks depends on the distribution of human capital (via $\Delta^m$), not the overall level of education. We proxy for $\Delta^m$ by examining the percentage difference between mean and median (pre-tax) household income, and compare these 'mean-median' gaps across countries and over time. Data on intergenerational income elasticity (Corak, 2006) offer a rough estimate for the differences in long run labor market mobility across countries. Although these measures are imperfect, they nonetheless offer an opportunity to examine the potential for protectionist surges through the lens of theory.

Figure 8 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. The leftmost panel charts the change in US real household mean and median incomes, indexed to 1974. The mean-median gap (defined as a percentage of median income) has risen steadily since the start of the period, roughly doubling over the course of 40 years. The panel on the right of Figure 8 then compares the mean-median gap across countries and over time.[36] It is clear from the figure that by this measure, inequality

[35] Indeed, if the potential trade war is sufficiently small, the median voter would 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.

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




Source: US Census; UK ONS; Eurostat

Figure 8: Income Gap Across Countries and Over Time

in the US and the UK has been rising systematically over the past four decades and is demonstrably higher than in other wealthy countries.

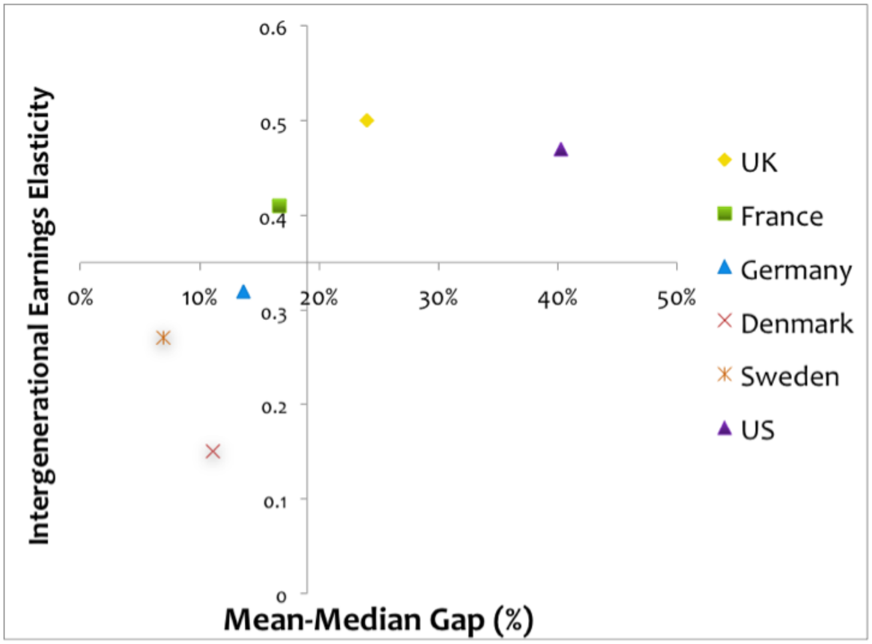
Figure 9 combines the most recent measures of the mean-median income gap for each country in Figure 8 with estimates of intergenerational income elasticity from Corak (2006). Higher values indicate less mobility between generations. We normalize the axes to the mean value of each indicator in the sample shown. The upshot is again immediate: the US and UK are both "stickier" and less equal than otherwise comparable developed countries.[37]

One critique of using these types of indicators is that they may not capture the potential for educational investments to lower inequality or raise social mobility. The data, however, suggest that education is no guarantee of higher individual earnings or falling overall inequality. According to recent work by Haskel, Lawrence, and Slaughter (2012), 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.[38] We replicate their figure with permission in Figure 10.

Overall, these data paint a picture of rising inequality and limited social mobility in the US and UK relative to other developed nations. Our theoretical exercise suggests

[37]This same pattern is seen in the famous "Great Gatsby Curve", introduced in a speech by Alan Krueger in 2012 when he was chair of the Council of Economic Advisors. Both depictions use the same elasticity estimates from Corak to measure social mobility, but Krueger used gini coefficients to measure inequality rather than the mean-median gap.

[38]See Blanchard and Willmann (2016) for static model of trade and endogenous educational polarization consistent with these findings.



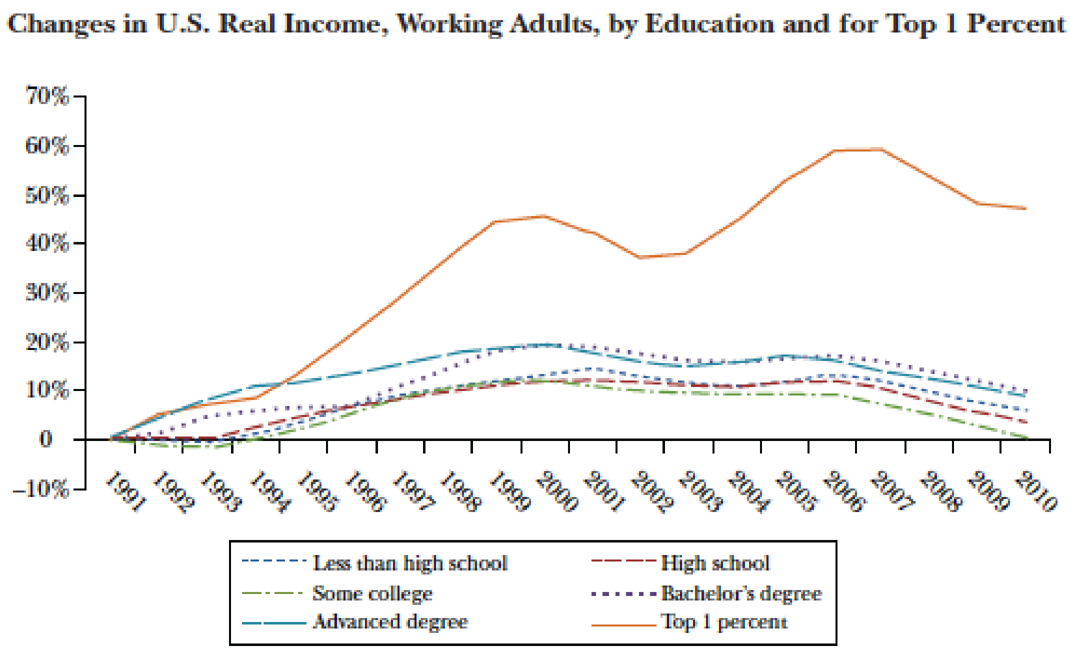

Source: Corak (2006); US Census; UK ONS; Eurostat

Figure 9: Inequality and Stickiness across Countries

that these economic factors may have played a role in the recent anti-globalization political shift in these countries. But do not take this too far. Economics is one of many drivers of electoral outcomes, and trade policy in particular played only a small role in the 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: increasing dispersion in economic prospects increases individuals' incentives to use policy 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.

## 6 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. 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 a surge




Source: US Census via Haskel, Lawrence, Slaughter *JEP* 2012

Figure 10: Education does not guarantee rising income

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 on the 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, which suggest that the US and UK are outliers relative to other OECD countries, with relatively low economic mobility and high inequality.

Our analysis raises a number of important questions for further research. Continuing in the context of trade and economic nationalism, our model could be extended to incorporate endogenous voter turn-out in response to macroeconomic shocks. Along a different line, one could explore the opportunity for inter-generational transfers to reduce or eliminate protectionist surges. In the long run, to what extent could a time-consistent Constitutional agreement or transfer scheme prevent overshooting or reverse escalation?

More broadly, 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 exercise 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: adoption of new technologies, political responses to climate change, pension reforms, and beyond.

## 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 Marketstates. *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, May). Untangling technology and trade: Evidence from local labor markets. *The Economic Journal 125*(584), 621–646.

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.

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 106(4). *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, 2014. *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.

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*.

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, September). Democratic public good provision. CEPR working paper No. 4044.

Hassler, J., K. Storesletten, and F. Zilibotti (2007, March). Democratic public good provision. *Journal of Economic Theory 133*(1), 127–51.

Jain, S. and S. Mukand (2003). Redistributive promises and the adoption of economic reform. *American Economic Review 93*(1), 256–264.

Krusell, P. and J. Ríos-Rull (1996). Vested interestes in a positive theory of stagnation and growth. *Review of Economic Studies 62*, 301–29.

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.

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.

Varian, H. (1992). *Microeconomic Analysis* (Third Edition ed.). WW Norton.

# A Supplementary Figures

The following figures offer a graphical supplement to the proofs of Propositions 4 and 5. Here, we work in $(\Delta_t^m, \tau_t)$ space to match the analytical proofs as closely as possible. Note that the $\tau_t(\Delta_t^m)$ locus is the same in and out of steady state, since all of the arguments of the tariff function are contemporary to time $t$. Conversely, for the out-of-steady-state function $\Delta_t^m = \Delta^m(\tau_{t-1}, \tau_t)$, the previous period's tariff acts as a shift variable. Thus, following the terms of trade shock, the equilibrium adjustment path is defined as the series $(\Delta_t^m, \tau_t)$ where for each period $t$, $\Delta_t^m = \Delta^m(\tau_{t-1}, \tau_t)$ intersects the (new) steady state tariff locus.

Figure A.1 illustrates the overshooting case. If $\Delta^{m\prime}(p) > 0$, then $\Delta^{m\prime}(\tau) < 0$; Assumption 2 ensures that the $\Delta^m(\tau)$ is steeper than the optimal tariff function and intersects $\tau(\Delta^m)$ only once and from above, as shown. Lemma 3 implies that the out-of-steady-state $\Delta^m(\tau_{t-1}, \tau_t)$ schedules are steeper than the steady state schedule, $\Delta^m(\tau)$. (Intuitively, $\Delta^m$ is less responsive in one argument than in both arguments together.)

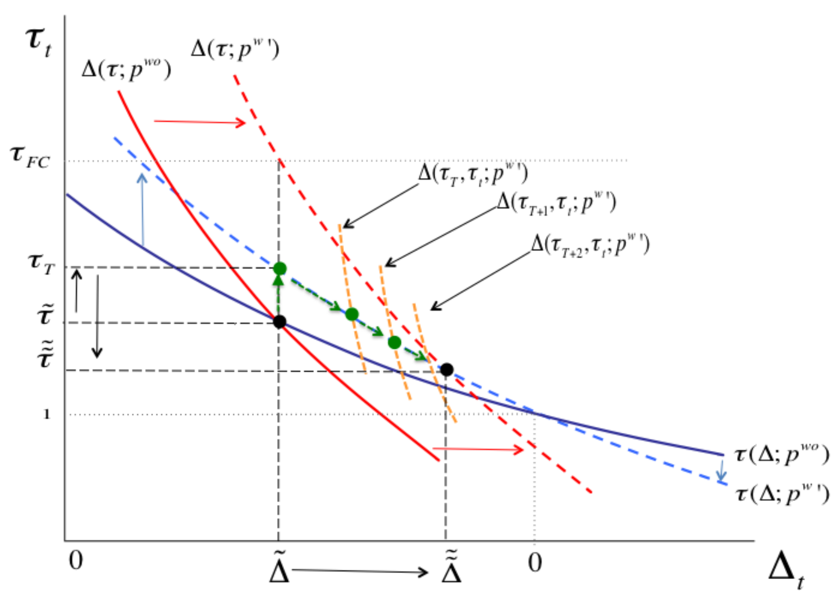



Figure A.1: Overshooting $(\Delta^m, \tau)$

Starting from a steady state $(\tilde{\Delta}^m, \tilde{\tau})$, and unanticipated increase in the terms of trade from $p^{wo}$ to $p^{w\prime} > p^{wo}$ at time $T$ causes the tariff locus to pivot around $\Delta^m = 0$, as shown. Under our assumption that $\tilde{\Delta}^m < 0$, and because the old generation cannot adjust their educational choices, (i.e. $\Delta^m_T = \Delta^m(\mathbf{e_{t-1}}) = \tilde{\Delta}^m$), the tariff jumps to $\tau_T$ (Lemma 3). The time $T+1$ (of of steady state) function is then given by $\Delta^m(\tau_T, \tau_t; p^{w\prime})$, which is steeper than the new steady state tariff locus, and intersects the tariff locus $\tau_t(\Delta^m_t; p^{w\prime}$ as shown. Thus, $\tau_{T+1} < \tau_T$ (and $\Delta^m_{T+1} > \Delta^m_T$). As the tariff falls, the $\Delta^m(\cdot,\cdot)$ schedule continues to shift right, as shown, converging to the new steady state.

We use the same logic to establish convergence for the case of protectionist escalation in Figure A.2, though the mechanics differ. In this case, since $\Delta^{m\prime}(p) \leq 0$, then $\Delta^{m\prime}(\tau) \geq 0$. Again, the out-of-steady-state $\Delta^m(\tau_{t-1}, \tau_t)$ schedules are steeper than the steady state schedule by Lemma 3, as shown.

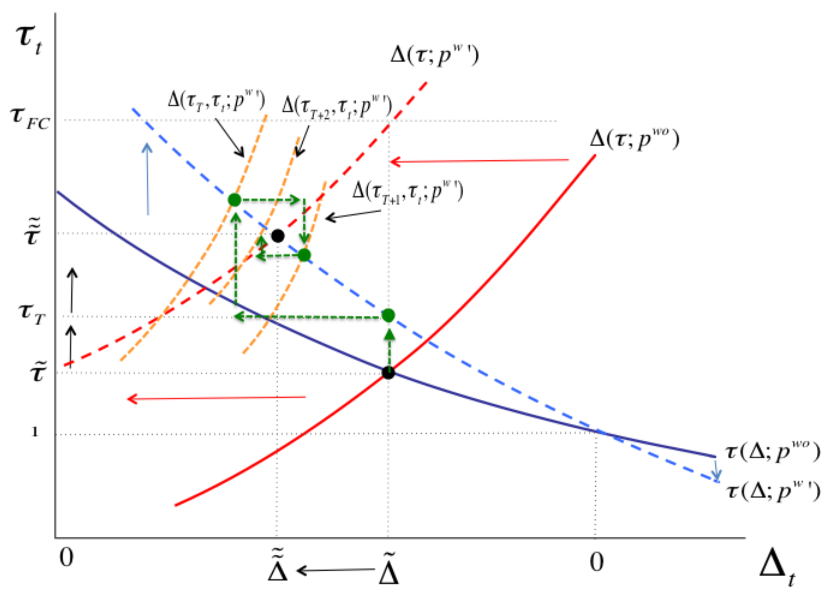



Figure A.2: Escalation in $(\Delta^m, \tau)$

In the escalation case, the terms of trade shock has the same effect on the tariff locus as in Figure A.1, but the opposite impact on the steady state $\Delta^m(\tau)$ schedule. (For any given tariff, higher $p^w$ implies a lower domestic price thus lower $\Delta^m$ since $\Delta^{m\prime}(p) \leq 0$.) As before, at the time of the shock, $\Delta^m_T = \tilde{\Delta}^m$, and so the tariff jumps to $\tau_T$. At $T+1$, the new equilibrium is given by the intersection of the new tariff schedule and $\Delta^m(\tau_T, \tau_t; p^{w\prime})$. Since $p_T < \tilde{p}$, we know that this schedule looks as shown, which results in $\tau_{T+1} > \tau_T$ and $\Delta^m_{T+1} < \Delta^m_T$. Because the time $T+1$ tariff is higher, the time $T+2$ $\Delta^m(\tau_{T+1}, \tau_t; p^{w\prime})$ shifts to the right, again as shown. Convergence proceeds by oscillation: when the tariff rises, inequality falls, which pushed the subsequent tariff lower; that 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.

# B Supplementary Proofs (Claims 1-6)

Below, we establish a series of useful properties that concern the optimal tariff expression, $\tau_t({\Delta^m}_t) = \tau_t(a^m, \mathbf{e_{t-1}})$ which is implicitly defined by the first order condition in (16). To streamline notation, we drop the time subscripts in the proofs below, with the recognition that the tariff at time $t$, $\tau_t$, depends on the contemporaneous values of $\Delta_t^m$ and $p_t^w$ and the lagged distribution of education $\mathbf{e}_{t-1}$. We use superscript $^o$ to denote the value of the optimal tariff given by the first order condition.

**Claim 1.** $V_{\tau\tau}(\tau, \Delta^m)\Big|_{\tau^o} < 0$

*Proof.* Recall the expression for $V_\tau$ from the first order condition in (16):

$$V_\tau \;\; = \;\; \nu(p)p\left[-\frac{\Delta^m}{\tau} + t\cdot\frac{d\bar{E}}{d\tau}\right]. \tag{B.1}$$

Taking the derivative with respect to $\tau$:

$$V_{\tau\tau} \;\; = \;\; \nu(p)p\left(\frac{\Delta^m}{\tau^2} + \frac{d\bar{E}}{d\tau} + t\frac{d^2E}{d\tau^2}\right),$$

where the second line uses the envelope condition and we use $\nu(p) = \nu_I$. Evaluated at the optimal tariff: $\frac{\Delta^m}{\tau^o} = t^o \cdot \frac{d\bar{E}}{d\tau}$ from the first order condition,

$$V_{\tau\tau}\Big|_{\tau^o} = \nu(p)p\left[\left(\frac{t+\tau}{\tau}\right)\cdot\frac{d\bar{E}}{d\tau} + t\cdot\frac{d^2\bar{E}}{d\tau^2}\right]. \tag{B.2}$$

The expression for total net exports is given implicitly by $E = \bar{x}^s - \bar{d}^s$. With Cobb-Douglass preferences, then: $\bar{E} = \bar{x}^s + \frac{\alpha}{p}(p\bar{x}^s + 1 + tp\bar{E})$. Thus, $\bar{E} = \frac{(1-\alpha)\bar{x} - \frac{\tau\alpha}{p^w}}{(1+t\alpha)}$. Taking the derivative with respect to $\tau$:

$$\begin{aligned}
\frac{dE}{d\tau} \;\; &= \;\; \frac{-\alpha/p^w}{(1+t\alpha)} - \alpha\left(\frac{(1-\alpha)\bar{x} - \tau^2/p^w}{(1+t\alpha)^2}\right) \\
&= \;\; \left(\frac{-\alpha/p^w}{(1+t\alpha)^2}\right)((1-\alpha)\bar{x}p^w - \tau\alpha + 1 + \alpha t) \\
&= \;\; \left(\frac{-\alpha/p^w}{(1+t\alpha)^2}\right)(1-\alpha)(\bar{x}p^w + 1) < 0.
\end{aligned} \tag{B.3}$$

Taking the derivative again,

$$\begin{aligned}\frac{d^2E}{d\tau^2} &= \underbrace{(1-\alpha)(\bar{x}p^w+1)\left(\frac{-\alpha/p^w}{(1+\alpha t)2}\right)}_{=\frac{dE}{d\tau}}\left(\frac{-2\alpha}{(1+\alpha t)}\right)\\ &= \frac{dE}{d\tau}\left(\frac{-2\alpha}{(1+\alpha t)}\right) > 0 \end{aligned} \tag{B.4}$$

Substituting (B.3) and (B.4) into (B.2) yields:

$$\begin{aligned} V_{\tau\tau}\Big|_{\tau^o} &= \nu(p)p\left(\frac{t+\tau}{\tau} - \frac{2\alpha t}{(1+\alpha t)}\right)\frac{dE}{d\tau}\\ &= \underbrace{\nu(p)p}_{+}\left[\underbrace{\left(\frac{1+t(1-\alpha)}{\tau(1+\alpha t)}\right)}_{+}\underbrace{\frac{dE}{d\tau}}_{-}\right] < 0. \end{aligned}$$

□

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

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

$$V_{\tau p^w}\Big|_{\tau^o} = \nu(p)p\Big(t^o\frac{\partial^2\bar{E}}{\partial\tau\partial p^w}\Big). \tag{B.5}$$

Then take the derivative of (B.3) with respect to $p^w$ to get:

$$\frac{\partial^2\bar{E}}{\partial\tau\partial p^w} = \frac{\alpha(1-\alpha)}{(1-\alpha t)^2}\frac{1}{p^{w2}} > 0 \tag{B.6}$$

Substituting (C.3) into (C.2) yields:

$$V_{\tau p^w} = \underbrace{\nu(p)p}_{+}\Big(t^o\underbrace{\frac{dE^2}{d\tau dp^w}}_{+}\Big).$$

Thus, the sign depends on the sign of the initially optimal tariff, $t^o$, which depends in turn on the sign of $\Delta^m$:

$$\begin{aligned}\Delta^m < 0,\ t^o > 0 \ \Rightarrow\ V_{\tau p^w} > 0\\ \Delta^m \geq 0;\ t^o \leq 0 \ \Rightarrow\ V_{\tau p^w} \leq 0.\end{aligned}$$

□

**Claim 3.** $V_{\tau a}(a,e)\Big|_{\tau^o} < 0$.

*Proof.* Taking the derivative of (B.1) with respect to $a$ (holding $e$ fixed), we have:

$$V_{\tau a} = \nu(p)p\left(-\frac{1}{\tau}\frac{\partial\Delta(a,e)}{\partial a}\right)$$

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

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

Substituting gives the result:

$$V_{\tau a}\Big|_{\tau=\tau^o} = \underbrace{\nu(p)p}_{+}\left(-\frac{1}{\tau}\right)\underbrace{\frac{\partial\Delta(a,e)}{\partial a}}_{+} < 0$$

□

**Claim 4.** $V_{\tau e}(a,e)\Big|_{\tau^o} < 0$

*Proof.* Taking the derivative of (B.1) with respect to $e$ (holding $a$ fixed), we have:

$$\begin{aligned} V_{\tau e} &= \nu(p)p\left(-\frac{1}{\tau}\frac{\partial\Delta(a,e)}{\partial e}\right) \\ \text{where, } \frac{\partial\Delta(a,e)}{\partial e} &= (1-\alpha)b\frac{\partial h(a,e)}{\partial e} > 0 \\ \text{s.t., } V_{\tau e}\Big|_{\tau^o} &= \underbrace{\nu(p)p}_{+}\left(-\frac{1}{\tau}\right)\underbrace{\left(\frac{\partial\Delta(a,e)}{\partial e}\right)}_{+} < 0 \end{aligned}$$

□

**Claim 5.** *The expression $t\tau\frac{dE}{d\tau}$ is decreasing in $\tau$.*

*Proof.*

$$\frac{d}{d\tau}\left(t\tau\frac{dE}{d\tau}\right) = \frac{dE}{d\tau}(t+\tau) + t\tau\frac{d^2E}{d\tau}.$$

Substituting from (B.4):

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

□

**Claim 6.** *Holding education fixed,* $V_{\tau b}|_{\tau^o} > 0$ *if and only if* $\Delta^m < 0$.

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

$$V_{\tau b} = \nu_I \frac{p}{\tau}\left\{t\tau\frac{\partial^2 E}{\partial\tau\partial b} - \frac{\partial\Delta^m(\mathbf{e},b)}{\partial b}\right\}. \tag{B.7}$$

Using $h(a^m) \equiv h(a^m, e(a^m))$ and $\bar{h} \equiv \int_a h(a, e(a))f(a)da$ as shorthand, we have:

$$\begin{aligned}\Delta^m &= (1-\alpha)b(h(a^m) - \bar{h}), \text{ which implies}\\ \frac{\partial\Delta^m(\mathbf{e},b)}{\partial b} &= (1-\alpha)(h^m - \bar{h}) = \frac{\Delta^{mo}}{b} = t\tau\left(\frac{1}{b}\right)\frac{d\bar{E}}{d\tau},\end{aligned}$$

where $\Delta^{mo}$ denotes the initial value of $\Delta^m$ and the last equality uses the first order condition for the initially optimal tariff, $\Delta^{mo} = t^o\tau^o\frac{dE}{d\tau}$.

Substituting into (B.7) and collecting terms, we have:

$$V_{\tau b} = \nu_I p t^o\left\{\frac{\partial^2 E}{\partial\tau\partial b} - \frac{1}{b}\frac{d\bar{E}}{d\tau}\right\}$$

Recall that $\frac{d\bar{E}}{d\tau} = \frac{-\alpha(1-\alpha)}{(1+\alpha t)^2}\left(b\bar{h} + \frac{1}{p^w}\right)$, so that:

$$\frac{\partial^2\bar{E}}{\partial\tau\partial b} = \frac{-\alpha(1-\alpha)}{(1+\alpha t)}\cdot\bar{h}$$

Substituting again:

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

□

# C Proof of Lemma 3

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

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

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 $\Delta^m$ functions.

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

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

At the same time,

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

Substituting (C.2) and (C.3) into (C.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:

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

Next, we establish that if the steady state is unique and stable, then the same must be true for the out of state equilibrium pairs, $(\Delta^m{}_t, \tau_t)\forall t$. To do this, it is sufficient to show that the out-of-steady-steady schedules $\Delta^m{}_t(p_t, p_{t+1})$ are less responsive in each argument than is $\Delta^m(p)$ in $p$. The following claim and corollary show this result:

**Claim 7.** *The steady state function $\Delta^m(p)$ is more responsive to changes in $p$ than is the out-of-steady-steady schedule $\Delta^m{}_t(p_t, p_{t+1})\forall t$ 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* $\frac{d\Delta^m(p)}{dp} \leq \frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_t}, \frac{\partial\Delta^m(p_t,p_{t+1})}{\partial p_{t+1}} \leq 0.$

*Proof.* From the definition in (16):

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

So,

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_t} = (1-\alpha)b\left[h_e\frac{\partial e(a^m; p_t, p_{t+1})}{\partial p_t} - \int h_e\frac{\partial e(a; p_t, p_{t+1})}{\partial p_t}f(a)da\right] \tag{C.4}$$

Likewise,

$$\frac{\partial\Delta^m(p_t, p_{t+1})}{\partial p_{t+1}} = (1-\alpha)b\left[h_e\frac{\partial e(a^m; p_t, p_{t+1})}{\partial p_{t+1}} - \int h_e\frac{\partial e(a; p_t, p_{t+1})}{\partial p_t}f(a)da\right] \tag{C.5}$$

From Lemma 1:

$$\frac{\partial e(a;p_t,p_{t+1})}{\partial p_t} = \frac{\alpha}{1-\alpha}\ \frac{p_{t+1}}{p_t}\ \frac{\partial e(a;p_t,p_{t+1})}{\partial p_{t+1}} \tag{C.6}$$

Combining (C.4)-(C.6) 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,

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

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

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

□

Claim 7 implies in turn that if $\Delta^m(\tau)$ is increasing in $\tau$, then the out-of-steady-state schedule, $\Delta^m(\tau_{t-1},\tau_t)$ will also be increasing in $\tau_t$: assuring that out-of-steady-state equilibrium is (also) stable and unique. Conversely, if $\Delta^m(\tau)$ is decreasing in $\tau$ faster than the tariff schedule, so that steady state is unique and stable, then the same will also be true out of steady state: $\Delta^m(\tau_{t-1},\tau_t)$ will be decreasing even faster in $\tau_t$ than the tariff schedule $\tau({\Delta^m}_t)$. Thus:

**Corollary 7.** *If the steady state equilibrium is unique and stable, so is the out-of-steady-state equilibrium in each period $t$.*

□