Population Dynamics of Welfare Stigma: Welfare Fraud vs. Incomplete Take-Up

Jun-ichi Itaya* Kenichi Kurita†

April 5, 2023

Abstract

There are two important problems in welfare benefit programs: the prevalence of *welfare fraud*, in which ineligible people receive welfare benefits, and *incomplete take-up*, whereby eligible poor people are reluctant to claim welfare benefits. This study investigates both of these opposing phenomena by extending the static models of statistical discrimination and the taxpayer resentment view welfare stigma suggested by Besley and Coate (1992) to a simple dynamic model of a population game. We find multiple stable equilibria in the long run, one of which entails *low welfare fraud and 100% incomplete take-up* and the other of which entails *high welfare fraud and complete take-up* in either model, and, moreover, that an interior stationary equilibrium that allows for the coexistence of welfare fraud and incomplete take-up is unstable in the first model, but it is stable in the second model. We thus conclude that the model of the tax-payer resentment view welfare stigma would provide a better explanation for the persistent coexistence of *welfare fraud* and *incomplete take-up* actually observed in Japan and Germany.

Keywords: Stigma, Population Game, Incomplete Take-up, Welfare Fraud, Non-Take-Up

JEL classifications: H31, H53, I38

*Corresponding author: School of Economics, Department of Law and Economics, Hokusei Gakuen University, 2-3-1, Ohyachi-Nishi, Atsubetsu-ku, Sapporo 004-8631, Japan. E-mail: j-itaya@hokusei.ac.jp

†Urban Institute and Department of Civil Engineering, Faculty of Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan, Tel: +81-92-802-3429, Email: kurita.kenichi.564@m.kyushu-u.ac.jp

1 Introduction

There are two important problems in welfare benefit programs: the prevalence of *incomplete take-up*, whereby eligible poor people are reluctant to claim welfare benefits, and *welfare fraud,* whereby ineligible poor people claim welfare benefits. This study investigates both problems in a simple framework of population dynamics with respect to heterogenous populations. In particular, we focus on how welfare stigma, which is a negative marking of households who take up welfare benefits, forms over time as a result of population dynamics. Although welfare stigma is important in understanding the emergence of welfare fraud and incomplete take-up, we find that the dynamic evolution of the composition between eligible (or deserving) claimants and ineligible (or undeserving) claimants also critically matters in determining welfare stigma.

The existence of welfare stigma means that some people choose not to receive public assistance despite satisfying the eligibility criteria. Table 1 shows take-up rates in Japan, the UK, the US, and Germany. The table reports that take-up rates are less than 100%, indicating incomplete take-up.[1]

	Data	Take up rate
Tachibanaki and Urakawa (2006)[JPN]	SRI	16.3∼19.7%
Duclos (1995)[UK]	FES	80%
Blank and Ruggles (1996)[US]	SIPP	60∼67%
Riphahn (2001)[GER]	EVS	37%

Table 1 (Adapted from Tachibanaki and Urakawa; 2006).

Table 1 also reveals an extremely lower take-up rate of eligible recipients

[1] Tachibanaki and Urakawa (2006) define the "take-up rate" as (the number of households taking up welfare)/(the number of eligible households) using the date from Survey on the Redistribution of Income (SRI), where eligible households are the ones with lower income than the poverty line. Duclos (1997) estimates the take-up rate as (the number of *eligible* households taking up welfare)/(the number of *eligible* household) by using the data from Family Expenditure Survey (FES). Blank and Ruggles (1996) estimate the take-up rate in the US using the data from Survey of Income and Program Participation (SIPP). They used two definitions for a take-up rate: (the number of households taking up welfare)/(the number of *eligible* household+the number of non-eligible households) and (the number of *eligible* households taking up welfare)/(the number of *eligible* household). Riphahn (2001), on the other hand, estimates the "non-take-up rate" as (the number of eligible households who do not take up welfare)/(the number of *eligible* household). Except for the definitions of Tachibanaki and Urakawa, and the first definition of Blank and Ruggles, the other definitions coincide with our definition regarding the variable p.

in Japan. According to Takahashi (2017), the key to understanding this phenomenon is the existence of welfare stigma because they fear negative labels, disapproval, or public shaming if they participate in a public assistance program. Though the public assistance system is designed to complement the social insurance system, the widely observed *low* take-up rate of public assistance in Japan is obviously problematic because it creates inefficiency thereby preventing the social security system from functioning as expected. This implies that the social security system is incomplete unless the public assistance system works properly. Hence, not only has welfare stigma been of academic interest in sociology and economics for the past several decades, but its reduction is also considered to be an important social policy issues in welfare programs.

On theoretical grounds, Moffitt (1983), Besley and Coate (1992, henceforth BS92), Yaniv (1997), and Blumkin et al. (2015) analyze a welfare stigma model focusing on welfare fraud. They find that stigma could be an alternative to law enforcement for suppressing welfare fraud. However, incomplete take-up is beyond the scope of these studies, except for Moffitt (1983), which allows for endogenous choices on whether or not to take up benefits, but not for welfare fraud. The present study aims to fill this research gap by considering together *welfare fraud* and *incomplete take-up* in a simple population dynamics framework.[2]

There are several innovations and findings in this study. First, we introduce endogenous choices of taking up welfare recipients into BS92's models of the *statistical discrimination view stigma*, which allows for endogenous choices of the *only* undeserving poor as to whether or not to become welfare fraud. This extension intends to clarify how incomplete take-up emerges endogenously, which BS92 and Blumkin et al. (2015) do not address. To do this, we need to allow for the populations of the deserving poor who qualify for welfare benefits and the undeserving poor who are **not** qualified to take up welfare benefits to simultaneously and endogenously change through time. More specifically, we employ a framework of population dynamics that endogenously and jointly determines the populations of both deserving and undeserving claimants through time.

Second, BS92 stipulate that the equilibrium level of welfare stigma is determined as a **fixed point** of their stigma cost function. More precisely, the level of stigma cost is determined by itself as a sort of rational expec-

[2]Kurita et al. (2020) investigate the effect of changes in benefit levels on a proportion of recipients to total population empirically as well as theoretically. Although the populations of deserving and non-deserving poor are treated as endogenous variables like the present model, their model is essentially a static model. Hence, their main focus lies on empirical studies.

tations equilibria in the sense that all individuals can precisely predict the stigma costs in equilibrium. Consequently, no further revisions to know the precise level of stigma costs take place at that equilibrium as a result of a sort of thought experiments. This scenario requires a great deal of common knowledge and god-like calculation powers in a timeless world. More importantly, it seems unclear to us why welfare claimants stick to finding a *stationary value* (or fixed point) of stigma costs, although this equilibrium concept plays a central role in comparative static analysis of BS92. Despite the fact that the explanatory significance of equilibrium concepts in general depends on the *plausibility* of the underlying dynamics that bring the players to equilibrium, their timeless scenario on how to reach this equilibrium is not especially convincing to describe the actual behavior of welfare beneficiaries.

In contrast, applying the dynamic model of population games (see, e.g., Sandholm, 2001) allow us to explain how the level of stigma costs is formed through changes in the *composition* between the deserving and undeserving claimants over time instead of the process of thought experiments in terms of stigma costs suggested by BS92. Moreover, our dynamic model is more compatible with the model of the *statistical discrimination view stigma* in the sense that the size of the stigma cost is obviously sensitive to the composition of population between the deserving and undeserving claimants; indeed, higher populations of the current welfare fraud induce more fraud, thereby increasing stigma costs.

Third, the present population dynamics give rise to multiple long run equilibria, which BS92 do not address. The multiplicity of long run equilibria are caused by aggregate population externalities arising from changes in the demographic composition of the heterogeneous poor through varying a stigma cost function. The existence of multiple equilibria undermines the predictive power of a comparative static analysis as well as the policy evaluation because considerably different comparative static results may emerge depending on which stationary equilibrium to choose. As a consequence, the comparative statics analysis carried out by BS92 may lead to misleading predictions because they focus only on a unique fixed point.

Fourth, and closely related to third point, our population dynamics model provides a method to refine the equilibrium, when multiple equilibria arises. The stability analysis of the population dynamics helps provide a way to refine the equilibria. Although the stability analysis reduces the number of equilibria in the *statistical discrimination view stigma* model to a large extent, we find that there may exist **at most two** asymptotically stable stationary equilibria: one of which entails welfare fraud to some extent, but completely eliminates incomplete take-up (i.e., all deserving poor individuals end up taking up welfare benefits), and the other of which allows for welfare fraud to some extent and 100% incomplete take-up (i.e., no deserving poor individuals take up welfare benefits). The comparative statics analysis focusing on these two stationary points reveals that the population of undeserving claimants rises in response to increasing welfare benefits, as well as reductions in the degree of public exposure and wage rates, while that of deserving claimants is **unaffected** by any parameter changes in the long run. This result provides another important policy implication: since the first equilibrium with full take-up is more socially desirable compared to the second one with 100% incomplete take-up, there is a room for governmental policy intervention in such a way that the policy of increasing welfare benefits would be socially desirable on the ground that this policy makes it more likely that the society will autonomously reach the second equilibrium with full take-up in the long run through population dynamics.

Fifth, we also construct a simple dynamic model of population games of *taxpayer resentment view stigma* suggested by *BS92.* Although there also exist multiple long run equilibria, we can use the stability analysis to pin down a **unique** stable long run equilibrium; one of the above-mentioned boundary stationary equilibria prevails if an interior stationary equilibrium does not exist. Conversely, if an interior stationary equilibrium exists, it is the only stable one, which allows for the persistent **coexistence** of welfare fraud and incomplete take-up in the long run.

The organization of the paper is as follows. The next section and Section 3 describe the basic model and then characterize the dynamic model of population games of the *statistical discrimination view stigma.* Sections 4 and 5 conduct a stability analysis and a comparative static analysis with respect to the structural parameters of the model as well as policy instruments, respectively, to address policy implications. Section 6 performs the same analysis in the dynamic model of population games of taxpayer resentment view stigma. Section 7 concludes the paper with a discussion of our findings and suggestions for future research questions. Some mathematical proofs are relegated to the appendices.

2 The model of *statistical discrimination view stigma*

Although we consider a society composed of the poor and the rich income classes, we here ignore the rich because they do not play any role in the present *statistical discrimination view stigma* model. We normalize the poor population to be equal to 1. The poor income class is further divided into two types: the deserving (*eligible*) poor and undeserving (*ineligible*) poor. The deserving poor are **unable** to work physically even if they want to work and are the intended targets of welfare benefits, while the undeserving poor are **able** to work if they want to do so. For analytical simplicity, the populations of the deserving and undeserving poor are fixed at γ and $1-\gamma$, respectively.

The government sets the benefit level at an *exogenously fixed* value of b, which may be equal to the minimum standard of living. The undeserving poor can get a *fixed* wage rate ω if they work, but suffers from disutility θ, while the deserving and undeserving poor both suffer from *stigma costs* when they receive welfare benefits. As a result, the payoffs to the deserving and undeserving poor are respectively displayed as follows:

$$U_{\text{deserving poor}} = \begin{cases} u(b) - s(p,q) & \text{if taking-up welfare benefits,} \\ 0 & \text{if not taking-up welfare benefits,} \end{cases} \tag{1}$$

$$U_{\text{undeserving poor}} \begin{cases} v(b) - s(p,q) & \text{if taking-up welfare benefits,} \\ v(\omega) - \theta & \text{if working,} \end{cases} \tag{2}$$

where $s(p,q)$ represents the stigma cost suffered when taking up benefits, which depends on the share of *actual* recipients in the population of deserving poor, denoted by p, and the share of *actual* recipients in the population of undeserving poor, denoted by q, and $u(b)$ and $v(b)$ respectively represent the utility levels of the deserving and undeserving poor, when taking-up welfare benefits, while $v(\omega)$ the utility level of the undeserving poor when working. Without loss of generality, we assume $\omega > b > 0$ (i.e., $v(\omega) > v(b)$), which guarantees that the undeserving poor are willing to work rather than to take up benefits if the disutility arising from labor supply measured by θ were to be relatively low. However, it is important to note that there are no *direct* strategic interactions through which an individual plays against an other individual. Instead, there are *aggregative* strategic interactions through which an individual plays against a whole society (more precisely, the same population to which that individual belongs). Since the latter externalities could be viewed as the population externalities caused by demographic changes in the deserving and undeserving poor, the present game would be called as "a *population game*" (see, e.g., Sandholm 2010).

We also assume that the degree of the disutility of work among the undeserving poor, θ, is increasing to 1 and uniformly distributed over the interval $[0, 1]$ as illustrated in the lower diagram of Fig.2. As a result, there exists an undeserving individual with $\hat{\theta}$ who is indifferent between working and not working; hence, given values of p and q, $\hat{\theta}$ is determined according to

$$v(\omega) - \hat{\theta} = v(b) - s(p,q). \tag{3}$$

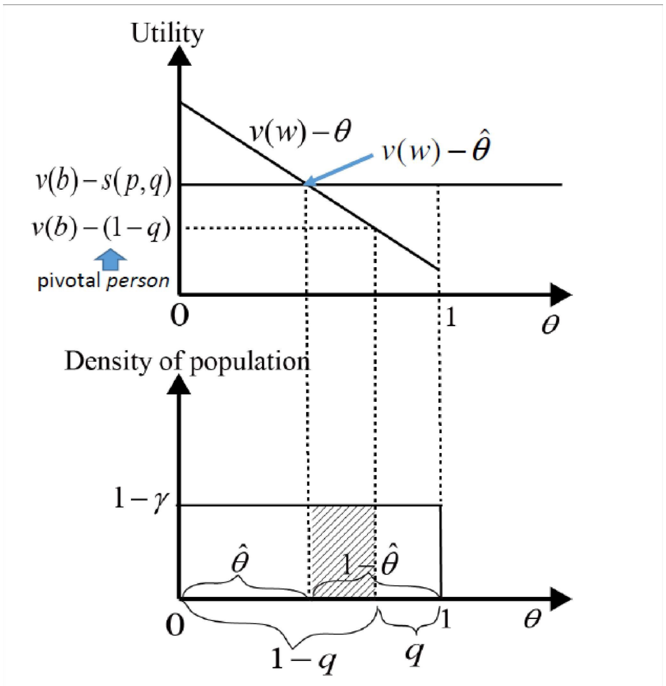

Figure 1: The shaded area corresponds to the population of the unsatisfied undeserving poor

All individuals in the undeserving poor group who have $\theta \geq \hat{\theta}$ prefer to take up welfare benefits over working; otherwise, prefer to work. However, since q is historically given in the present model at each moment in time, the population of *actual* undeserving recipients, q, may not coincide with that of undeserving poor who want to take up benefits, $1 - \hat{\theta}$. If they do not coincide each other, the discrepancy between q and $1 - \hat{\theta}$ represents the *unsatisfied* undeserving poor group who want to take up benefits but does not take up them, which is illustrated as the shaded area in the lower diagram of Fig.1. Since the undeserving poor individual whose utility is given by $v(b) - (1 - q)$ (whom we call a **pivotal undeserving individual)** is most frustrated by the largest discrepancy, she will be the first person to change the strategy of working to that of taking up benefits. Nevertheless, all unsatisfied undeserving individuals do not immediately change their strategies because some of them can change their strategies benefits following the *proportional imitation rule* (see, e.g., Hofbauer and Sigmund, 1998) under the imitation dynamics described in Section 3.

Following BS92, we also adopt a stigma cost function which is an increasing function of the discrepancy between the average disutility of work among all actual welfare claimants, denoted by $\bar{\theta}_w$, and the average disutility of work among the poor, denoted by $\bar{\theta} = \int_0^1 \theta d\theta = 1/2$. To obtain closed-form solutions for q and p, we assume that the stigma cost function is a *linear* function of the difference $\bar{\theta}_w - \bar{\theta}$:[3]

$$s(\bar{\theta}_w - \bar{\theta}) = \lambda[\bar{\theta}_w - \bar{\theta}], \tag{4}$$

where the constant parameter λ measures the degree of public exposure in welfare programs.[4] **The average disutility among all actual welfare claimants** $\bar{\theta}_w$ is given by

$$\bar{\theta}_w = \pi\bar{\theta}_d + (1-\pi)\bar{\theta}_u, \tag{5}$$

where π is the fraction of deserving claimants:

$$\pi = \frac{\gamma p}{\gamma p + (1-\gamma)q}, \tag{6}$$

$\bar{\theta}_d$ is the average disutility of work among the deserving claimants, and $\bar{\theta}_u$ is the average disutility of work among the undeserving claimants:

$$\bar{\theta}_u = \int_{1-q}^{1} \frac{\theta d\theta}{q} = 1 - \frac{q}{2}. \tag{7}$$

For simplicity, we assume that $\bar{\theta}_d = \bar{\theta} = 1/2$, implying that $\bar{\theta}_u \geq \bar{\theta}_d$ since $q \leq 1$.

Using (5), (6) and (7), together with $\bar{\theta}_d = \bar{\theta} = 1/2$, we can express the difference in (4) by

$$\bar{\theta}_w - \bar{\theta} = (1-\pi)\left(\bar{\theta}_u - \frac{1}{2}\right) = \frac{1}{2}\frac{(1-\gamma)q(1-q)}{\gamma p + (1-\gamma)q}. \tag{8}$$

[3]We also adopt the Besley and Coate (1992)'s assumption in that the stigma is the function of the difference between the mean labor disutility in the welfare recipients and that in the society in the statistical discrimination view model. The reason is that the idea is consistent with the property of stigma suggested by Goffman (1963) as follows:

"... it was suggested that a discrepancy may exist between an individual's virtual and actual identity. This discrepancy, when known about or apparent, spoils his social identity; it has the effect of cutting him off from society and from himself so that he stands a discredited person facing an unaccepting world." (Line 19-23, Page. 19, Goffman, 1963)

[4]The size of λ would be determined by institutionary and social psychological factors. For example, when recipients use foodstamps, they may deeply feel embarrassed in public, meaning higher λ. The strict *means test* imposes public exposure to a large extent. In Japan, for example, applicants are required to exhaust all or most of their savings, and relatives whom the authorities view as possible sources of support were also required to submit to means tests, meaning higher public exposure and thus higher λ.

By substituting (8) into (4), we can rewrite the stigma cost function (4) as follows:

$$s(p,q) = \frac{\lambda}{2}\frac{(1-\gamma)q(1-q)}{\gamma p + (1-\gamma)q}. \tag{9}$$

As (9) makes clear, the stigma cost is affected not only by the parameters b, ω, λ, and γ, but also by the time-varying state variables p and q. Note, however, that point $(p,q) = (0,0)$ should not be contained in the domain of (p,q) because $s(p,q)$ is *discontinuous* at $(0,0)$.

3 Imitation Dynamics

To explain how the equilibrium population state is actually reached in our population game, we need to model the behavior of the deserving and undeserving poor individuals at the micro-level. To do this, we consider the model of imitation dynamics (see, e.g., Hofbauer and Sigmund, 1998) in which at each moment in time an individual samples another individual from the same population and she mimics the strategy adopted by the sampled individual in the same population who obtained a payoff at least as high as herself; otherwise, she stays with the same strategy as before.

In the model of imitation dynamics just outlined, suppose that the population proportion of deserving poor individuals p have taken up benefits at time t. They currently earn $u(b) - s(p,q)$, and randomly sample from the same population. At time t, a fraction p^2 of the deserving poor who have taken up benefits samples another deserving poor individuals who have taken up benefits. They do not switch their strategies. A fraction $p\,(1-p)$ of the deserving poor who have taken up benefits samples another deserving poor individuals who have **not** taken up benefits with the payoff 0, while the same fraction of the deserving poor who have **not** taken up benefits samples another deserving poor individuals who have taken up benefits with the payoff $u(b) - s(p,q)$. They may or may not swith depending on whether $u(b) - s(p,q) \lesseqgtr 0$. A fraction $(1-p)^2$ of the deserving poor who have **not** taken up benefits samples another deserving poor individuals who have **not** taken up benefits with the payoff 0. They do not switch their strategies. With the proportional imitation rule (Schlag, 1994), which says imitate actions that perform better with a probability proportional to the difference in the expected gains, the growth rate of deserving claimants is given by

$$\begin{aligned} \dot{p}/p &= \left[u(b) - s(p,q) - (u(b) - s(p,q))\left(p^2 + p(1-p)\right) - 0 \cdot \left((1-p)p + (1-p)^2\right)\right] \\ &= (1-p)\left[u(b) - s(p,q)\right]. \end{aligned} \tag{10}$$

The story is similar for the population of undeserving poor individuals. Suppose that the population proportion of undeserving poor individuals, q, has taken up benefits at time t. They earn $v(b) - s(p,q)$, and randomly sample from the same population. The undeserving poor individuals switch away from their current strategy only if the payoff associated with the strategy of taking up benefits, $v(b) - s(p,q)$, is **less** than that associated with the strategy of working, $v(w) - \theta$. As before, the growth rate of undeserving claimants is given by

$$\begin{aligned} \dot{q}/q &= [v(b) - s(p,q)] - [q\,(v(b) - s(p,q)) + (v(\omega) - \theta)\,(1-q)] \\ &= (1-q)\,[v(b) - s(p,q) - v(\omega) + (1-\theta)]\,. \end{aligned} \tag{11}$$

Although the parameter θ in (11) is distributed over the undeserving poor, the value of θ must be pinned down into a unique value to depict the dynamics of q. We set $\theta = 1-q$. This is because if q were to be in the position depicted in Fig. 1, the pivotal individual with $\theta = 1-q$ who initially worked finds by sampling that $v(b) - s\,(p,q) > v(w) - (1-q)$, and she is willing to switch to take up benefits. Then, the current pivotal individual will be replaced by a new pivotal one with lower disutility of work, and thus she is also willing to do so; consequently, such replacement continues to occur (i.e., $\dot{q} > 0$) until q is equal to q^* such that $v(b) - s\,(p,q^*) = v(w) - (1-q^*)$ holds; namely, the unsatisfied undeserving poor ends up being completely eliminated in the long run, given p. The growth rate of undeserving claimants (11) is thus expressed by

$$\dot{q}/q = (1-q)\,[v(b) - s(p,q) - v(\omega) + (1-q)]\,. \tag{12}$$

Next, we identify stationary equilibria. To do this, we set $\dot{p} = 0$ in (10) and $\dot{q} = 0$ in (12) within the unit square $[0,1] \times [0,1]$ of R^2_+ (except for $(0,0)$), which yields $p^* = 0$ or 1, $q^* = 0$ or 1,

$$u(b) - s(p^*,q^*) = 0, \tag{13}$$

$$\text{and } v(b) - v(\omega) + 1 - q^* - s(p^*,q^*) = 0, \tag{14}$$

where p^* and q^* denote the stationary values of p and q, respectively. To simplify the analysis, we further make the following assumptions:

Assumption 1 $1 > (2u(b)/\lambda)$.

Assumption 2 $\lambda < 2$.

An immediate consequence of Assumptions 1 and 2 is $u(b) < 1$. Although not all combinations of those values do not exist as a stationary equilibrium under Assumptions 1 and 2, we can prove the following lemma (the proof is given in Appendix A):

Lemma 1 *Under Assumptions 1 and 2 there exist only stationary equilibria for the system* (10) *and* (12) *such that*

$$(p^*, q^*) = \{(1, \bar{q}), (\hat{p}, \hat{q}), (0, \tilde{q}), (0, \tilde{q}_2), (1, 0), (0, 1), (1, 1)\}, \tag{15}$$

where $\tilde{q}$ and $\bar{q}$ are respectively the intersection points between (14) *and the axis $p = 0$, as well as the axis $p = 1$, $\tilde{q}_2$ is the intersection point between* (13) *and the axis $q = 0$, and $(\hat{p}, \hat{q})$ is the interior intersection point between* (13) *and* (14).

4 Stability

In this section, we investigate the stability properties of the stationary points of the system (10) and (12). Although there are many stationary points, we will focus on only *stable* stationary points in order to perform a meaningful comparative statics analysis.

For the stability analysis, we take a linear approximation of (10) and (12) around the stationary point (p^*, q^*):

$$\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} \partial f(p^*, q^*)/\partial p & \partial f(p^*, q^*)/\partial q \\ \partial g(p^*, q^*)/\partial p & \partial g(p^*, q^*)/\partial q \end{bmatrix} \begin{bmatrix} p^* - p \\ q^* - q \end{bmatrix}, \tag{16}$$

where the functions $f(\cdot)$ and $g(\cdot)$, respectively, represent the right-hand sides of (10) and (12), and

$$\frac{\partial f(p^*, q^*)}{\partial p} = (1 - 2p^*)\left[u(b) - s(p^*, q^*)\right] - p^*(1 - p^*)s_p(p^*, q^*),$$

$$\frac{\partial f(p^*, q^*)}{\partial q} = -p^*(1 - p^*)s_q(p^*, q^*), \quad \frac{\partial g(p^*, q^*)}{\partial p} = -q^*(1 - q^*)s_p(p^*, q^*), \text{ and}$$

$$\frac{\partial g(p^*, q^*)}{\partial q} = (1 - 2q^*)\left[v(b) - v(\omega) + 1 - q^* - s(p^*, q^*)\right] - q^*(1 - q^*)\left[1 + s_q(p^*, q^*)\right],$$

noting that the stationary equilibrium values of p^* and q^* are given by (15).

Differentiating (9) with respect to p and q, respectively, yields

$$s_p(p, q) \equiv \frac{\partial s(p, q)}{\partial p} = -\frac{\lambda(1 - \gamma)}{2} \frac{\gamma q(1 - q)}{\left[\gamma p + (1 - \gamma)q\right]^2} < 0, \text{ and} \tag{17}$$

$$s_q(p, q) \equiv \frac{\partial s(p, q)}{\partial q} = \frac{\lambda(1 - \gamma)}{2} \frac{\gamma p(1 - 2q) - (1 - \gamma)q^2}{\left[\gamma p + (1 - \gamma)q\right]^2} \gtreqless 0. \tag{18}$$

On the stability properties of the system (10) and (12) around the respective stationary points, we can demonstrate the following proposition:

Proposition 1 *Under the population dynamics* (10) *and* (12) *coupled with Assumptions 1 and 2 for all* $(p, q) \in [0, 1]^2 \setminus (0, 0)$*:*

(i) The stationary point $(\hat{p}, \hat{q})$ *is a saddle.*

(ii) The stationary point $(1, \bar{q})$ *is locally asymptotically stable (LAS) if* $u(b) - s(1, \bar{q}) > 0$*. Conversely, if* $u(b) - s(1, \bar{q}) < 0$*, then it is a saddle.*

(iii) The stationary point $(0, \tilde{q})$ *is LAS if* $u(b) - s(0, \tilde{q}) < 0$*. Conversely, if* $u(b) - s(0, \tilde{q}) > 0$*, then it is a saddle.*

(iv) The stationary point $(1, 0)$ *is LAS if* $v(b) - v(w) + 1 < 0$*, while if* $v(b) - v(w) + 1 > 0$*, then it is a saddle. The stationary point* $(1, 1)$ *is a saddle and* $(0, 1)$ *is a source.*

(v) The stationary point $(0, \tilde{q}_2)$ *is **not** LAS.*

Proof. See Appendix B.[5] ■

Five remarks are in order. First, although there are multiple stationary equilibria as indicated by (15) in Lemma 1, the only stability analysis can help pin down fewer stationary points. It follows from Proposition 1 that the stationary equilibria such as $(1, \bar{q})$, $(0, \tilde{q})$ and $(1, 0)$ are stable provided the respective conditions stipulated in Proposition 1 are satisfied, as we see in Figs. 2-4.

Second, there may exist only one *stable* stationary equilibrium as illustrated in Figs. 2 and 3. If it is a case, the system (10) and (12) will reach a unique long run equilibrium; that is, either $(1, \bar{q})$ or $(0, \tilde{q})$ is globally asymptotically stable, independently of initial conditions. We call the former the *Besley and Coate equilibrium* (or simply, the *B&C* one) and the latter the *non-take-up equilibrium*. The latter equilibrium has not been identified in BS92, because they have eliminated the case of incomplete take-up by setting $p = 1$.

Insert Figure 2

Insert Figure 3

[5]It is straightforward to show that the eigenvalues of the Jacobian at every stationary point for the respective linearized systems are all real numbers and thus the trajectories never circle around the stationary point, provided that the stationary equilibrium is hyperbolic.

Third, the stationary equilibria $(1, \bar{q})$ and $(0, \tilde{q})$ are *simultaneously* stable if an interior stationary equilibrium exists, as illustrated in Fig.4. In other words, since the interior stationary equilibrium $(\hat{p}, \hat{q})$ is a saddle, it is a threshold in the sense that it separates the basins of attraction of the stationary equilibria $(1, \bar{q})$ and $(0, \tilde{q})$. This feature implies that the trajectories of p and q starting from different initial conditions could converge to different stationary points in the long run. If the initial population of p is relatively larger, then the trajectory starting from such an initial condition is more likely to lead to the Besley and Coate equilibrium $(1, \bar{q})$. Intuitively, since the stigma cost is lower due to the higher initial p (recalling (17)), more of the deserving poor are willing to take up welfare benefits. Hence, the higher p leads to a further reduction in the stigma cost, and thus this process is repeated through the imitation dynamics. Due to such a self-enforcing cumulative process of imitation, the population of deserving claimants p keeps rising through time according to (10) until *all* deserving poor end up taking up welfare benefits (i.e., $p = 1$ in the long run).

Insert Figure 4

Conversely, if the initial population of p is relatively smaller, then the stigma cost is higher due to (17), so that the above process is reversed. As a result, the population of deserving claimants p keeps falling through time until all deserving poor end up *giving up* benefits (i.e., $p = 0$ in the long run). The important point is that the self-enforcing cumulative process stated above ultimately leads to one of the boundary stationary equilibria (i.e., $(0, \tilde{q})$ or $(1, \bar{q})$) rather than the interior stationary equilibrium $(\hat{p}, \hat{q})$. This self-enforcing property stems from the *monotonically decreasing* stigma cost function with respect to the population of deserving claimants p (see (17)).[6] In contrast, since the effect of changes in q on the stigma cost is ambiguous in general (see (18)), the relation between the population of undeserving claimants q and the stigma cost is **not monotonic**; consequently, the movement of q does not display the self-enforcing property. As a result, q would reach an interior value (i.e., $\tilde{q}, \bar{q} \in (0, 1)$) in the long run, rather than 0 or 1. Moreover, the emergence of different stable stationary points entails different comparative statics properties. Hence, we should not expect a unique prediction concerning the long run comparative statics effects with respect to changes in the structural parameters of the model or the policy in-

[6]If the stigma cost function in (4) is concave in $\bar{\theta}_w - \bar{\theta}$, then this process becomes milder, while if the stigma cost function in (4) is convex in $\bar{\theta}_w - \bar{\theta}$, then this process will accelerate. In either case, the divergent tendency towards the boundary stationary points remains valid for even more general stigma cost functions.

struments, which makes a sharp contrast with the *unique* comparative statics result of BS92.

Fifth, our stability analysis reveals that the *interior* stationary point $(\hat{p}, \hat{q})$ is a saddle and thus unstable due to the above-mentioned cumulative divergence property, as illustrated in Fig.4. This striking and somewhat surprising result does not allow for the **coexistence** of welfare fraud and incomplete take-up in a long run. However, it seems that this extreme theoretical prediction does **not** align with actual observations, such as in Japan, Germany, and the like (see Table 1 also).

5 Comparative statics

Based on the stability results obtained in the previous section, we can confine our attention to only the *stable* long run equilibria, such as $(0, \tilde{q})$ and $(1, \bar{q})$, to make a meaningful comparative static analysis.

To sum up (the detailed deviations are relegated to Appendix C),

Proposition 2 *Under Assumptions 1 and 2:*

(i) An increase in the level of welfare benefits raises the population of undeserving claimants in both the B&C and non-take-up equilibria.

(ii) An increase in the degree of public exposure reduces the population of undeserving claimants in both the B&C and non-take-np equilibria.

(iii) An increase in the wage rate reduces the population of undeserving claimants in both the B&C and non-take-up equilibria.

(iv) An increase in the share of deserving poor in the total population of the poor raises the population of undeserving claimants in the B&C equilibrium, but reduces that in the non-take-up equilibrium.

(v) No parameter change affects the population of deserving claimants in both the B&C and non-take-up equilibria.

Let us first suppose that in the stable *non-take-up* equilibrium $(0, \tilde{q})$, the level of welfare benefits increases by a small amount of $\Delta b > 0$. The increased welfare benefit induces more of both deserving and undeserving poor to take up the benefits because their utilities of receiving the benefits (i.e., $u(b)$ and $v(b)$) are higher than those of **not** receiving. Since the stable stationary point $(0, \tilde{q})$ must be located inside the area where $\dot{p} = p(1-p)\left[u(b) - s(0, \tilde{q})\right] < 0$ due to Proposition 1 (iii), as illustrated in Figs. 3 and 4, the increased

$u(b+\Delta b)$ would mitigate the downward pressure on $\dot{p}$, but $\dot{p} < 0$ remains for arbitrarily small $\Delta b > 0$. In spite of $\dot{p} < 0$, p must stay at 0 because p is initially set equal to its lower limit 0. At the same time, because the increased b also lowers $\hat{\theta} = v(\omega) - v(b+\Delta b) + s(0, \tilde{q})$, more of the undeserving claimants with *lower disutility of work* are willing to take up benefits rather than to work, thereby raising $q = 1 - \hat{\theta}$. The rise in q reduces the average disutility of undeserving claimants $\theta_u = 1 - (q/2)$, and so does $\bar{\theta}_w = \pi\bar{\theta}_d + (1-\pi)\bar{\theta}_u = \bar{\theta}_u$ (recalling $\pi = 0$ due to $p = 0$). As a result, the discrepancy $\bar{\theta}_w - \bar{\theta}$ falls (recalling $\bar{\theta} = (1/2)$), as does the stigma cost (which is also consistent with $s_q(0, \tilde{q}) < 0$ in (18)). The decreased stigma cost further reduces $\hat{\theta}$ and thus further raises q. Since this same process is repeated indefinitely, q keeps rising until the equality $v(b) - v(\omega) + 1 - \tilde{q}' - s(0, \tilde{q}') = 0$ with $q^* > \tilde{q}$ in (14) is resorted. Hence q ends up with being higher in a new stationary equilibrium $(0, \tilde{q}')$.

Next, consider the effect of an increase in b in $(1, \bar{q})$. Since $(1, \bar{q})$ is a stable stationary point, p must stay at 1 in spite of the increased b for the similar reason as before. Since the increase in b also reduces $\hat{\theta} = v(\omega) - v(b + \Delta b) + s(1, \bar{q})$, more of the undeserving claimants with *lower disutility of work* are willing to take up benefits rather than to work, thereby raising $q = 1 - \hat{\theta}$. However, the final effect on $\hat{\theta}$ may be uncertain, because the increased q also affects the stigma cost but its effect is ambiguous (recalling $s_q(1, \tilde{q}) \gtreqless 0$ in (18)). As indicated by (C.2) in Appendix C, nevertheless, the **direct** effect of increasing b on the stigma cost dominates the ambiguous **indirect** effect of the induced change in q on the stigma cost; consequently, q rises.

Consider an increase of λ in the stable *non-take-up* equilibrium $(0, \tilde{q})$. Although an increase in λ *directly* raises the stigma cost (recall (9)), p must stay at 0. Furthermore, $\hat{\theta} = v(\omega) - v(b) + s(0, \tilde{q})$ also rises due to the increased stigma cost. This implies that less of the undeserving claimants are willing to take up benefits rather than to work, thereby lowering $q = 1 - \hat{\theta}$. The decreased q raises the average disutility of undeserving claimants $\theta_u = 1 - (q/2)$ in (7) and so does $\bar{\theta}_w = \pi\bar{\theta}_d + (1-\pi)\bar{\theta}_u = \bar{\theta}_u$ (noting that $\pi = 0$ due to $p = 0$). As a result, the stigma cost is further increased (which is also confirmed by $s_q(0, \tilde{q}) < 0$ in (18)), thereby further lowering $q = 1 - \hat{\theta}$. Since these two effects together raise the stigma cost, so does $\hat{\theta}$; consequently, q falls in the new *non-take-up* equilibrium.

In the stable B&C equilibrium $(1, \bar{q})$ the increase in λ also directly raises the stigma cost thereby raising $\hat{\theta}$ and thus lowering q as before. In addition, the indirect effect through changes in q on the stigma cost may be uncertain as before. As stated above, the direct *negative* effect of λ on the stigma cost dominates the *ambiguous* indirect effect through varying q. As a result, the stigma cost unambiguously rises and thus $q = 1 - \hat{\theta}$ falls (see (C.3) in Appendix C).

BS92 focus on the comparative statics effects on the stigma cost rather than the populations of deserving and undeserving claimants, and find that the effect of b on stigma costs is **ambiguous** in general because a rise in benefits not only raises the fraction of undeserving claimants, $1 - \pi$, but also induces more undeserving claimants to take up benefits thereby reducing $\bar{\theta}_u$. Due to these two conflicting effects on the average disutility $\bar{\theta}_w = \pi\bar{\theta}_d + (1 - \pi)\bar{\theta}_u$, the final effects on $\bar{\theta}_w$ and thus on the stigma cost are uncertain in general. Although the effect on the stigma cost may be uncertain at $(1, \bar{q})$ in our model also (recalling $s_q(1, \bar{q}) \gtreqless 0$ due to (18)), we see that **the direct effects of changes in the parameters such as b and λ on the stigma cost dominate the indirect effect through changes in q**. Consequently, those effects on q will be unambiguously determined in this model.

There are noteworthy policy implications that differ from those in BS92. BS92 are solely concerned with how to raise the stigma cost in order to reduce *welfare fraud*, while we are chiefly concerned with how to improve lower or zero take-up rate, because policymakers in countries such as Germany and Japan may put a more weight on the population of deserving claimants as policy targets rather than the level of stigma costs. Since the welfare of deserving claimants p remains unaffected in the long run, one may tend to deduce the conclusion in which the policy instrument λ would be socially more desirable than b on the grounds that the increased λ reduces the population of undeserving claimants (i.e., welfare fraud), but the increased b does not. However, based on our population dynamics model, we would provide different policy implications. If the society is more concerned with improving incomplete take-up among the deserving poor, there is a room for governmental policy intervention that makes the long run equilibrium $(1, \bar{q})$ more likely to occur.[7] In implementing this policy prescription, the government has to shift the locus $\dot{p} = 0$ *downwards* while shifting the locus $\dot{q} = 0$ *upwards* in order to make the stationary point $(1, \bar{q})$ more likely to be stable by expanding the basin of attraction of the stationary equilibrium $(1, \bar{q})$. To this end, we need to lower the intercept of the locus $\dot{p} = 0$ with the q-axis, denoted by $q^*|_{\dot{p}=0}$, as well as to raise that of the locus $\dot{q} = 0$ with the q-axis,

[7]Note, however, that since the population of undeserving claimants $\bar{q}$ in $(1, \bar{q})$ would be larger than $\tilde{q}$ in $(0, \tilde{q})$, it is a trade-off relation between the welfare of the undeserving poor and the number of welfare fraud. Nevertheless, it seems that the problem of *non-take-up* in $(0, \tilde{q})$ is more serious for societies such as Japan.

denoted by $\tilde{q}|_{\dot{q}=0}$, where

$$q^*|_{\dot{p}=0} = 1 - \frac{2u(b)}{\lambda} \text{ and } \tilde{q}|_{\dot{q}=0} = 1 - \frac{2\left[v(\omega) - v(b)\right]}{2-\lambda}. \tag{19}$$

Inspecting (19) reveals that higher levels of b reduce the intercept $q^*|_{\dot{p}=0}$, while raising the intercept $\tilde{q}|_{\dot{q}=0}$, both of which make it more likely to realize Fig. 2 rather than Fig. 3. Concluding, we can say that raising the level of welfare benefits is more socially desirable than enhancing the degree of public exposure provided the society is more concerned with the welfare of deserving poor.

6 The model of taxpayer resentment view stigma

In this section, we introduce an alternative stigma costs function based on the taxpayer resentment view of welfare stigma suggested by BS92 into the population dynamics model. We begin by briefly outlining their model. Although they postulate not only that the level of welfare benefits b is exogenously chosen by the government, but also that the cost of providing the benefits b is financed by the lump-sum taxes, T, borne **only** by the rich. The population of the rich is $1-\beta$, while the population of the poor is $\beta \in (0,1)$. Thus, we can express the government's budget constraint as:

$$(1-\beta)T = b\beta\left[\gamma p + (1-\gamma)q\right]. \tag{20}$$

The left-hand side represents the total tax revenue borne by the rich, while the right-hand side represents the cost for providing the total welfare benefits. Thus, the level of the lump-sum tax must be equal to

$$T = \frac{b\beta\left[\gamma p + (1-\gamma)q\right]}{1-\beta}, \tag{21}$$

with the following properties:

$$\frac{\partial T}{\partial p} = \frac{b\beta\gamma}{1-\beta} > 0 \text{ and } \frac{\partial T}{\partial q} = \frac{b\beta(1-\gamma)}{1-\beta} > 0. \tag{22}$$

We assume that the rich individual's income, y, is constant over time, common across the rich population, and satisfies

Assumption 3 $y - T > 0$.

Following BS92, we assume not only that the preferences of the rich are additively separable in private consumption, $y - T$, and the concerns (or degrees of compassion) regarding the aggregate consumption or poverty of the deserving poor; more specifically, the rich have different concerns μ about the degree of poverty $P(b)$ with $P'(b) < 0$.[8] Each rich individual characterized by a given value of μ chooses the **most** preferred benefit level, denoted by $b^*(\mu; p, q)$, by maximizing his/her utility function, as follows:

$$\begin{aligned} b^*(\mu; p, q) &= \arg\max_{\{b\}} \{\log[y - T] - \mu\beta\gamma P(b)\}, \\ &= \arg\max_{\{b\}} \left\{ \log\left[y - \frac{b\beta[\gamma p + (1-\gamma)q]}{1-\beta} \right] - \mu\beta\gamma P(b) \right\}, \end{aligned}$$

where the second equality follows from substituting for T in (21). Note that the rich can *see-through* the government's budget constraint such that each rich individual chooses b^* *subject to the government's budget constraint* (20).

The first-order condition for maximization with respect to b is

$$\left[y - \frac{\beta(\gamma p + (1-\gamma)q)}{1-\beta} b \right]^{-1} \frac{\beta[\gamma p + (1-\gamma)q]}{1-\beta} = -\mu\beta\gamma P'(b). \tag{23}$$

The left-hand side of (23) represents the marginal utility cost of an increase in one unit of the tax payment, while the right-hand side represents the marginal benefit of improving the well-being of the poor. To obtain a closed-form solution for b, we further assume $P(b) = -\eta b + c$ with a constant $\eta > 0$, whereby (23) simplifies to

$$\frac{\gamma p + (1-\gamma)q}{(1-\beta)(y-T)} = \mu\gamma\eta. \tag{24}$$

Recalling (21), we can solve (24) for b to obtain

$$b^*(\mu; p, q) = \frac{(1-\beta)y}{\beta[\gamma p + (1-\gamma)q]} - \frac{1}{\beta\gamma\eta\mu}. \tag{25}$$

Since the rich individuals (i.e., taxpayers) have different values for μ, they have *different most preferred benefit levels*. Nevertheless, there certainly exists a critical individual having the *threshold concern* $\bar{\mu}$ such that his/her most preferred benefit level $b^*(\bar{\mu}; p, q)$ precisely coincides with the benefit level *exogenously* chosen by the government, b.

$$b^*(\bar{\mu}; p, q) = b. \tag{26}$$

[8] This assumption implies that rich individuals believe that an increase in welfare benefits mitigates poverty.

As a result, the remaining taxpayers are decomposed into two groups: those whose μ are larger than $\bar{\mu}$ will regard the prevailing level of welfare benefits as being too parsimonious, while those whose μ are less than $\bar{\mu}$ view it as being too generous. Solving (26) for $\bar{\mu}$, coupled with (25), yields

$$\bar{\mu} = \frac{\gamma p + (1-\gamma) q}{\gamma\eta(1-\beta)(y-T)}. \tag{27}$$

Let $r(\cdot)$ represent the **resentment** felt by each rich individual who regards the welfare benefit as excessive. For analytical simplicity, we also assume that the function $r(\cdot)$ is a linear and increasing function of the discrepancy between the actual benefits level b and the level that a rich individual with μ regards as being the most favorable (or appropriate) :

$$r(\mu; p, q) = \lambda^r \left[b - b^*(\mu; p, q) \right],$$

where the scale parameter of the stigma cost λ^r takes a constant and common value among the rich. As in BS92, we define the stigma costs function as an increasing function of *aggregate* taxpayers' resentment, such as

$$s^r(p,q) = (1-\beta)\int_0^{\bar{\mu}} r(\mu; p, q)\mu d\mu = (1-\beta)\int_0^{\bar{\mu}} \lambda^r \left[b - b^*(\mu; p, q) \right] \mu d\mu,$$

where we assume that μ is *uniformly distributed* over the interval $[0, 1]$. Then, using the government's budget constraint (20), we can rewrite the above stigma cost function as follows:

$$s^r(p,q) = (1-\beta)\lambda^r \left[\frac{b}{2}\bar{\mu}^2 - \int_0^{\bar{\mu}} \left\{ \frac{(1-\beta)\, y}{\beta \left[\gamma p + (1-\gamma) q \right]}\mu - \frac{1}{\beta\gamma\eta} \right\} d\mu \right].$$

Substituting (27) into $\bar{\mu}$ in the above expression yields

$$s^r(p,q) = \frac{\lambda^r}{2\beta\gamma^2\eta^2} \frac{\gamma p + (1-\gamma) q}{y - T}, \tag{28}$$

with the property

$$\frac{\partial s^r(p,q)}{\partial T} = \frac{\lambda^r}{2\beta\gamma^2\eta^2} \frac{\gamma p + (1-\gamma) q}{(y-T)^2} > 0. \tag{29}$$

This feature is quite intuitive in the sense that the increased tax burden borne by the rich T enhances their resentment, thereby raising the stigma cost incurred by all welfare benefit claimants.

6.1 Stability properties of the model of taxpayer Resentment View Stigma

The population dynamics is also described by (10) and (12), except that the stigma cost function $s(p,q)$ is replaced by $s^r(p,q)$ in (28). As in the previous model, the loci $u(b)-s^r(p,q)=0$ and $v(b)-v(\omega)+1-q-s^r(p,q)=0$ play a crucial role in drawing the phase portraits of p and q (see Figs. 6-9). The following lemma serves in drawing these two loci (the proofs are relegated to Appendix D):

Lemma 2 *Under Assumption 3*

(i) The slope of the locus $u(b)-s^r(p,q)=0$ is larger in absolute value than that of the locus $v(b)-v(\omega)+1-q-s^r(p,q)=0$ for $(p,q)\in[0,1]^2$.

(ii) The intercepts of the locus $v(b)-v(\omega)+1-q-s^r(p,q)=0$ on the vertical axis $p=1$, denoted by q^r, and on the vertical axis $p=0$, denoted by $\tilde{q}^r$, must be less than 1.

With Lemma 2, we can identify the stability properties of the respective stationary points as follows:

Proposition 3 *Under the population dynamics* (10) *and* (12) *with being replaced by the stigma cost function* (28)*, coupled with Assumptions 1, 2, and 3, for all $(p,q)\in[0,1]^2$:*

(i) The interior stationary point $(\hat{p}^r,\hat{q}^r)$ is LAS.

(ii) The stationary point $(1,\bar{q}^r)$ is LAS if $u(b)-s^r(1,\bar{q}^r)>0$. Conversely, if $u(b)-s^r(1,\bar{q}^r)<0$, then it is a saddle.

(iii) The stationary point $(0,\tilde{q}^r)$ is LAS if $u(b)-s^r(0,\tilde{q}^r)<0$. Conversely, if $u(b)-s^r(0,\tilde{q}^r)>0$, then it is a saddle.

(iv) The stationary point $(\bar{p}^r_2,1)$ is a saddle.

(v) The stationary point $(\tilde{p}^r_2,0)$ is LAS if $v(b)-v(\omega)+1-s^r(\tilde{p}^r_2,0)<0$. Conversely, if $v(b)-v(\omega)+1-s^r(\tilde{p}^r_2,0)>0$, then it is a saddle.

(vi) The stationary points $(1,1)$, $(0,0)$, and $(0,1)$ are either a source or a saddle.

(vii) The stationary point $(1,0)$ is LAS if $u(b)-s^r(1,0)>0$ and $v(b)-v(\omega)+1-s^r(1,0)<0$. Otherwise, it is a source or a saddle.

(viii) The stationary points $(1, \bar{q}_2^r)$, $(0, \tilde{q}_2^r)$*, and* $(\tilde{p}^r, 0)$ *are* ***not*** *LAS.*

Proof. See Appendix E. ■

Based on Lemma 2 and Proposition 3, Figs. 6-9 provide several phase portraits for the dynamic behavior of p and q.

Several remarks are in order. First, when there exists an *interior* stationary point such as $(\hat{p}^r, \hat{q}^r)$ in Fig. 6, it is LAS, while all other non-interior stationary points are unstable. As a result, $(\hat{p}^r, \hat{q}^r)$ is globally asymptotically stable. This noteworthy feature stands in sharp contrast with the previous statistical discrimination view stigma model displaying that the interior stationary point is *unstable*. The reason for this difference is as follows. In the taxpayer resentment model, if the initial populations of both deserving and undeserving claimants are too large (i.e., close to $(1, 1)$), then the rich bear a heavier taxpayment burden. Consequently, their resentment will be greatly intensified, thereby augmenting the stigma cost incurred by all claimants. Therefore, the resulting higher stigma cost *discourages* both deserving and undeserving claimants to take up welfare benefits. This in turn decreases the populations of both claimants thereby preventing them from approaching $(1, 1)$. On the contrary, when both claimants are very much lower (i.e., close to $(0, 0)$), then so is the stigma cost, thereby inducing more of both deserving and undeserving poor to take up welfare benefits, and thus preventing them from approaching $(0, 0)$. These properties indicate that there is always a force that leads all claimants to depart from the boundary stationary points $(0, 0)$ and $(1, 1)$.

Insert Figure 5

Second, *when there does not exist an interior stationary point,* we obtain Figs.6-9. First, it turns out that the stationary equilibrium $(1, 0)$ is stable provided $u(b) > s^r(1, 0) > v(b) - v(\omega) + 1$ (i.e., the deserving claimants have relatively stronger preferences towards welfare benefits than the undeserving claimants). In this case the population of deserving claimants p tends to rise since $u(b) > s^r(1, 0)$. The increased p augments tax payments and thus intensifies the resentment of the tax payers, which in turn augments stigma costs. Hence, $v(\omega) - 1 > v(b) - s^r(1, 0)$ remains valid, which causes downward pressure on q. This implies that $(1, 0)$ will be stable in Fig.6, as stated in Claim (*vi*). In contrast, since the stationary equilibrium $(0, 1)$ is always located above the locus $\dot{q} = 0$ due to Lemma 2 (ii), $\dot{q} < 0$ always holds, as illustrated in Figs.5-9. As a result, q always departs from 1, and thus $(0, 1)$ is unstable (see Proposition 3 (vi)).

Insert Figure 6

Third, if $u(b) > s^r(1, \bar{q}^r)$ holds, then $(1, \bar{q}^r)$ is stable because p is rising to 1 around a local neighborhood of $(1, \bar{q}^r)$ regardless of whether q may be rising or falling, as illustrated in Fig.7. Conversely, if $u(b) < s^r(0, \tilde{q}^r)$ holds, then $(0, \tilde{q}^r)$ is stable because p is falling to 0 around a local neighborhood of $(0, \tilde{q}^r)$, regardless of whether q may be rising or falling, as illustrated in Fig.8. Finally, if $v(b) - v(\omega) + 1 - s^r(\tilde{p}_2^r, 0) < 0$, q is falling to 0 around a local neighborhood of $(\tilde{p}_2^r, 0)$. Hence, $(\tilde{p}_2^r, 0)$ is stable, regardless of whether p may be rising or falling, as illustrated in Fig.9.

Insert Figure 7

Insert Figure 8

Insert Figure 9

6.2 Comparative statics for the model of taxpayer resentment view stigma

This subsection conducts a comparative static analysis of the taxpayer resentment view stigma model in terms of changes in the parameters b, λ^r, y, ω, and η, respectively. For that purpose, we need to focus our analysis on four stable stationary points: $(\hat{p}^r, \hat{q}^r)$ in Fig. 5, $(1, \bar{q}^r)$ in Fig. 7, $(0, \tilde{q}^r)$ in Fig. 8, and $(\tilde{p}_2^r, 0)$ in Fig. 9. We summarize the results as follows (the detailed mathematical derivations are relegated to Appendix F):

Proposition 4 *Under Assumptions 3:*

(i) The effect of an increase in the benefit level on the population of undeserving claimants is ***ambiguous*** *in the stationary equilibria $(\hat{p}^r, \hat{q}^r)$, $(1, \bar{q}^r)$, and $(0, \tilde{q}^r)$, while the effect on that of deserving claimants is also* ***ambiguous*** *in $(\hat{p}^r, \hat{q}^r)$ and $(\tilde{p}_2^r, 0)$.*

(ii) The effect of an increase in the degree of public exposure on the population of undeserving claimants is ***negative*** *in $(1, \bar{q}^r)$ and $(0, \tilde{q}^r)$ but has* ***no effect*** *on that of undeserving claimants in $(\hat{p}^r, \hat{q}^r)$, while the effect on that of deserving claimants is* ***negative*** *in $(\hat{p}^r, \hat{q}^r)$ and $(\tilde{p}_2^r, 0)$.*

(iii) The effect of an increase in the income level of the rich on the population of undeserving claimants is ***positive*** *in $(1, \bar{q}^r)$ and $(0, \tilde{q}^r)$ but has* ***no effect*** *on that of undeserving claimants in $(\hat{p}^r, \hat{q}^r)$, while the effect on that of deserving claimants is* ***positive*** *in $(\hat{p}^r, \hat{q}^r)$ and $(\tilde{p}_2^r, 0)$.*

(iv) *The effect of an increase in the wage rate on the population of undeserving claimants is* ***negative*** *in* $(\hat{p}^r, \hat{q}^r)$, $(1, \bar{q}^r)$, *and* $(0, \tilde{q}^r)$, *but has* ***no effect*** *on that of deserving claimants in* $(\tilde{p}_2^r, 0)$, *while the effect on that of deserving claimants is* ***positive*** *in* $(\hat{p}^r, \hat{q}^r)$.

(v) *The effects of an increase in the efficacy of social welfare programs on the population of undeserving claimants are* ***positive*** *in* $(1, \bar{q}^r)$ *and* $(0, \tilde{q}^r)$, *but there is* ***no effect*** *on that of undeserving claimants in* $(\hat{p}^r, \hat{q}^r)$, *while the effect on that of deserving claimants is* ***positive*** *in* $(\hat{p}^r, \hat{q}^r)$ *and* $(\tilde{p}_2^r, 0)$.

Let us first suppose that the level of welfare benefits increases by a small amount of Δb in the *interior* stationary point $(\hat{p}^r, \hat{q}^r)$. The increased welfare benefit, $b + \Delta b$, induces more of both deserving and undeserving poor to take up benefits over time, which is implied by (10) and (12). As a result, the total tax burden unambiguously increases, which intensifies the resentment of the taxpayers (i.e., the rich), thereby increasing the stigma cost. The increased stigma cost in turn causes both deserving and undeserving claimants to fall, thereby reducing the stigma cost. Due to these conflicting effects on the stigma cost, therefore, the net effect on q may be *ambiguous.* In contrast, an increase in λ^r directly **raises** the stigma cost, thereby reducing both deserving and undeserving claimants. The resulting decreases in p and q in turn **reduce** the stigma cost. Although these two effects counteract each other, the *direct positive* effect of higher λ^r on the stigma cost will dominate the *indirect negative* effects of the induced changes in p and q, which is implied by (F.5) in Appendix F. As a result, the stigma cost unambiguously rises, thereby further decreasing p and q, except in $(\hat{p}^r, \hat{q}^r)$. Since (F.3) in Appendix F does not contain λ^r, q is unaffected by changes in λ^r in $(\hat{p}^r, \hat{q}^r)$.

An increase in the average income level of the rich, y, on the other hand, makes the rich more generous toward welfare benefits claimants, thereby depressing the stigma cost. As a result, the populations of both deserving and undeserving claimants tend to rise over time except for the effect on $\hat{q}^r$ in $(\hat{p}^r, \hat{q}^r)$. Since the stigma cost must be equal to $u(b)$ in the long run, the stationary value of $\hat{q}^r$ in $(\hat{p}^r, \hat{q}^r)$ is determined solely by (F.3) in Appendix F, which implies that $\hat{q}^r$ is independent of y in $(\hat{p}^r, \hat{q}^r)$. In contrast, an increase in the wage rate, ω, depresses the population growth rate of undeserving claimants (12) because the increased wage rate encourages them to work more rather than taking up benefits. Note, however, that the population of deserving poor $\tilde{p}_2^r$ in $(\tilde{p}_2^r, 0)$ is unaffected by changes in ω, because changes in ω do not affect (F.1) in Appendix F. Finally, enhancing the efficacy of social welfare programs (i.e., increasing μ) in general raises the population growth rates of both deserving and undeserving claimants. This occurs because the enhanced efficacy makes the rich more generous, which in turn reduces the stigma cost, thereby raising the populations of both deserving and undeserving claimants. However, the population of undeserving claimants in the interior stationary equilibrium $(\hat{p}^r, \hat{q}^r)$ remains unaffected because (F.3) in Appendix F is independent of changes in μ.

7 Concluding Remarks

We investigated the case in which welfare fraud and incomplete take-up *simultaneously* emerge in the long run, two phenomena of apparently opposite nature, in a simple setting of population dynamics, allowing the *heterogeneous* populations of both deserving and undeserving claimants to endogenously evolve over time, unlike the *static* model of BS92. The present study found multiple stationary equilibria in the population dynamics models of the statistical-discrimination view stigma as well as the taxpayer resentment view stigma suggested by BS92. We applied the stability analysis to reduce substantially the set (or multiplicity) of possible long run equilibria, allowing us to sharpen the predictive power of the comparative statics analysis by ruling out unstable stationary equilibria. Nevertheless, there may coexist **at most two** stable stationary equilibria, both of which are boundary equilibria, in the statistical discrimination view stigma model. In this case, depending on the initial population shares of deserving and undeserving claimants, different long run outcomes will emerge; in other words, different stationary equilibria deliver different comparative statics results and thus different policy implications. Furthermore, *there is no **stable** interior stationary equilibrium in the statistical discrimination view stigma model*; instead, either of the boundary stationary equilibria with 100% incomplete take-up and with 100% take-up exclusively arises.

In contrast, in the taxpayer resentment view stigma model, the interior stationary equilibrium is globally asymptotically stable, provided it exists; consequently, irrespective of the initial population shares of deserving and undeserving claimants, the coexistence of those claimants will be present in the long run. This most significant difference between these two models stems from the different nature of their stigma cost functions. The stigma cost function assumed in the statistical-discrimination view stigma model depends critically on the **composition of the deserving and undeserving claimants,** whereas the stigma cost function assumed in the taxpayer resentment view stigma model relies on the **total population of claimants**. In the former model, *if the population of welfare fraud is initially large,* then the deserving poor who want to receive welfare benefits are discouraged by higher stigma costs. Ultimately, all deserving poor end up giving up benefits in the long run. On the contrary, *if the initial level of welfare fraud population is small*, then all deserving poor end up taking up benefits in the long run. These extremely different long run outcomes stem from the self-enhancing cumulative feature of the stigma cost function in terms of the deserving claimants, thus monotonically leading to either 100% incomplete take-up or 100% complete take-up in the long run. Put differently, this cumulative destabilizing feature stems from the fact that the average disutility of all welfare claimants always depends **negatively** on the population of deserving poor. Hence, if there are currently many deserving claimants, their sense of the gilt (i.e., stigma costs) diminishes and they are more likely to take up welfare benefits and vice versa. This causes the population of deserving claimants to either monotonically decrease towards 100% incomplete take-up or monotonically increase towards 100% take-up. In contrast, the taxpayer resentment model displays a relatively more stabilizing tendency leading to an interior stationary equilibrium that allows for the persistent (rather than transitory) coexistence of welfare fraud and incomplete take-up if the interior stationary point exists. This tendency stems from aggregate externalities arising from the **total population** of welfare claimants through the total tax payments, rather than the composition of heterogeneous populations. Since the latter model generates the persistent coexistence of welfare fraud and incomplete take-up, it would better fit actual observations in Japan, Germany, and the like, compared to the statistical discrimination model.

The model presented in this paper can be developed further in several directions. In particular, the results of our analysis rely critically on the restrictive structure of the present model. For example, the utility function is *separable* in private consumption and the disutility of work, the labor supply is *inelastic*, the stigma cost function is *linear*, as well as the population share between the deserving and undeserving poor and between the poor and the rich are both fixed over time. To make the model more realistic and to derive more robust results, it would be desirable to conduct an analysis under a more general model. Although we have assumed exogenously fixed levels of welfare benefits and/or lump-sum taxes imposed on the rich, this scheme is obviously unrealistic because welfare benefits are mainly financed by distortionary taxes in most countries. By introducing distortionary (possibly, non-linear) taxes, the government has to face incentive compatibility constraints to cope with the accompanying asymmetric information problem, as in Blumkin et al. (2015). Therefore, an optimal welfare program solution would be part of *non-linear* optimal income taxation, as in the Mirrlees (1971) framework, which is augmented to simultaneously allow for asymmetric information on taxpayers' earning ability and for the extended margin choices of the undeserving poor. Furthermore, from the political economy perspective, voters could choose the policy parameters *endogenously*, as Lindbeck et al. (1999) suggest. In this view, it would be quite interesting and important to incorporate the political aspect of endogenous policy choices into our population dynamics framework.

The analysis in this paper is a first step towards addressing these more realistic and interesting extensions of a welfare stigma analysis.

References

Besley, T. and Coate, S. (1992). Understanding welfare stigma: Taxpayer resentment and statistical discrimination. *Journal of Public Economics* 48:165–183.

Blank, R.B. and Ruglles, P. (1996). When do women use aid to families with dependent children and food stamps?: The dynamic of eligibility versus participation. *The Journal of Human Resources*, 31 (1): 57-89.

Blumkin, T., Margalioth, Y. and Sadka, E. (2015). Welfare stigma reexamined. *Journal of Public Economic Theory*, 17(6), 874-886.

Duclos, J. Y. (1995). Modelling the take-up of state support. *Journal of Public Economics*, 58(3), 391-415.

Goffman. E. (1963). *Stigma: Notes on the Management of Spoiled Identity*, Englewood Cliffs, Prentice-Hall Inc.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge.

Kurita, K., Hori, N. and Katafuchi, Y. (2020). Stigma model of welfare fraud and non-take-up: Theory and evidence from OECD panel data, *International Journal of Economic Theory*, 1-29.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2), 175-208.

Moffitt, R. (1983). An economic model of welfare stigma. *American Economic Review*, 73(5):1023–1035.

Lindbeck, S., Nyberg, A.S., and Weibull, J.W. (1999). Social norms and economic incentives in the welfare state, *Quarterly Journal of Economics*, 114(1), 1-35.

Riphahn, R. T. (2010). Rational poverty or poor rationality? The take-up of social assistance benefits. *Review of Income and Wealth*, 47(3), 379-398.

Schlag, K. (1994). Why imitate, and if so, how? Exploring a model of social learning, Mimeo, D.P.B-296, Universität *Bonn*.

Tachibanaki, T. and Urakawa, K. (2006). *A Study on Poverty in Japan*, Tokyo University Press.

Takahashi, M. (2017). Spatial probit analysis on welfare stigma: Evidence from Japan, DP2016-006, Panel Data Research Center at Keio University.

Yaniv, G. (1997). Welfare fraud and welfare stigma. *Journal of Economic Psychology*, 18(4), 435-451.

Appendix A: Proof of Lemma 1

We solve $u(b) - s(p^*, q^*) = 0$ (i.e., the locus $\dot{p} = 0$) for p, together with (9), to obtain

$$p^* = \frac{1-\gamma}{\gamma} q^* \left[\frac{\lambda(1-q^*)}{2u(b)} - 1 \right], \tag{A.1}$$

whose right-hand side is a quadratic function of q (which is typically illustrated by the locus $\dot{p} = 0$ in Figs. 2-4). This graph crosses the origin $(p^*, q^*) = (0, 0)$ and the q-axis at $\tilde{q}_2 = -(2u(b)/\lambda) + 1 \in (0, 1)$. Hence, it never crosses the line $q = 1$. Next, evaluating $v(b) - v(\omega) + 1 - q - s(p, q) = 0$ at $q = 1$ yields $v(b) - v(\omega) - s(p, 1) < 0$, indicating that the locus $\dot{q} = 0$ never crosses the line $q = 1$. Similarly, evaluating the above equation at $p = 1$ yields $v(b) - v(\omega) + 1 - q - s(1, q) = 0$. These facts imply that when the locus $\dot{q} = 0$ crosses the line $p = 1$, q must be less than 1. We can also rewrite (9) as

$$p^* = \frac{1-\gamma}{\gamma} q^* \left[\frac{\lambda(1-q^*)}{2\left[v(b) - v(\omega) + 1 - q^*\right]} - 1 \right], \tag{A.2}$$

whose right-hand side is a rational function of q (see the locus $\dot{q} = 0$ in Figs. 2-4). This locus $\dot{q} = 0$ crosses the origin $(p, q) = (0, 0)$ and the q-axis at

$$\tilde{q} = 2\frac{v(b) - v(\omega)}{2 - \lambda} + 1 \in (-\infty, 1).$$

Taken together, the loci $\dot{p} = 0$ and $\dot{q} = 0$ both never cross the line $q = 1$. Finally, when the intersection between the loci $\dot{p} = 0$ and $\dot{q} = 0$ exists, equalizing (A.1) with (A.2) yields

$$\frac{1-\gamma}{\gamma}q^*\left[\frac{\lambda(1-q^*)}{2u(b)}-1\right]=\frac{1-\gamma}{\gamma}q^*\left[\frac{\lambda(1-q^*)}{2\left[v(b)-v(\omega)+1-q^*\right]}-1\right],$$

which simplifies to $u(b)=v(b)-v(\omega)+1-q^*$, i.e., $q^*=v(b)-v(\omega)+1-u(b)$. Hence, only when $0<v(b)-v(\omega)+1-u(b)<1$, $0<q^*<1$ is established.

Appendix B: Proofs of Proposition 1

(i) Since the determinant evaluated at $(\hat{p},\hat{q})$ is given by

$$\begin{vmatrix} -\hat{p}(1-\hat{p})s_p(\hat{p},\hat{q}) & -\hat{p}(1-\hat{p})s_q(\hat{p},\hat{q}) \\ -\hat{q}(1-\hat{q})s_p(\hat{p},\hat{q}) & -\hat{q}(1-\hat{q})\left[1+s_q(\hat{p},\hat{q})\right] \end{vmatrix}=\hat{p}(1-\hat{p})\hat{q}(1-\hat{q})s_p(\hat{p},\hat{q})<0,$$

where $s_p(\hat{p},\hat{q})<0$ from (17), it is a saddle.

(ii) The determinant of the Jacobian for the linearized system (16), which is evaluated at $(1,\bar{q})$, is given by

$$\begin{vmatrix} -[u(b)-s(1,\bar{q})] & 0 \\ -\tilde{q}(1-\tilde{q})s_p(1,\bar{q}) & -\bar{q}(1-\bar{q})\left[1+s_q(1,\bar{q})\right] \end{vmatrix}=[u(b)-s(1,\bar{q})]\bar{q}(1-\bar{q})\left[1+s_q(1,\bar{q})\right],$$

where it follows from (18) that

$$1+s_q(1,\bar{q})=\frac{\gamma^2+\frac{\lambda}{2}(1-\gamma)\gamma+(1-\gamma)\bar{q}(1-\frac{\lambda}{2})\left[2\gamma+(1-\gamma)\bar{q}\right]}{\left[\gamma+(1-\gamma)\bar{q}\right]^2}>0.$$

Moreover, if $u(b)-s(1,\bar{q})>0$, then the determinant is positive and the trace of the Jacobian is negative, implying that $(1,\tilde{q})$ is locally asymptotically stable (LAS). Conversely, if $u(b)-s(1,\bar{q})<0$, then the determinant is negative and thus $(1,\bar{q})$ is a saddle.

(iii) The determinant evaluated at $(0,\tilde{q})$ is given by

$$\begin{vmatrix} u(b)-s(0,\tilde{q}) & 0 \\ -\tilde{q}(1-\tilde{q})s_p(0,\tilde{q}) & -\tilde{q}(1-\tilde{q})\left[1+s_q(0,\tilde{q})\right] \end{vmatrix}=-\left[u(b)-s(0,\tilde{q})\right]\tilde{q}(1-\tilde{q})\left[1+s_q(0,\tilde{q})\right],$$

where $1+s_q(0,\tilde{q})=1-(\lambda/2)>0$ from (18). Moreover, if $u(b)-s(0,\tilde{q})>0$, then the determinant is negative, and thus $(0,\tilde{q})$ is a saddle. Conversely, if $u(b)-s(0,\tilde{q})<0$, the determinant is positive and the trace is negative, implying that $(0,\tilde{q})$ is LAS.

(iv) The determinants evaluated at $(1,1)$, $(0,1)$, and $(1,0)$ are, respectively, given by

$$\begin{vmatrix} -u(b) & 0 \\ 0 & v(\omega) - v(b) \end{vmatrix} = -u(b)\left[v(\omega) - v(b)\right] < 0, \tag{B.1}$$

$$\begin{vmatrix} u(b) & 0 \\ 0 & -\left[v(b) - v(\omega)\right] \end{vmatrix} = u(b)\left[v(\omega) - v(b)\right] > 0, \text{ and} \tag{B.2}$$

$$\begin{vmatrix} -u(b) & 0 \\ 0 & v(b) - v(\omega) + 1 \end{vmatrix} = -u(b)\left[v(b) - v(\omega) + 1\right] \gtreqless 0. \tag{B.3}$$

The negative sign of (B.1) implies that $(1,1)$ is a saddle, while since the determinant (B.2) is positive and the trace is positive, $(0,1)$ is unstable. If $v(b) - v(\omega) + 1 < 0$, the determinant (B. 3) is positive and the trace is negative, implying that $(1,0)$ is *LAS*. If $v(b) - v(\omega) + 1 > 0$, (B. 3) is negative, and thus $(1,0)$ is a saddle.

(v) Since the determinant evaluated at $(0, \tilde{q}_2)$ is equal to 0, $(0, \tilde{q}_2)$ is *non-hyperbolic*. Hence, we cannot apply the linear approximation technique. Nevertheless, it can easily be verified that one eigenvalue of its Jacobian is positive, and thus $(0, \bar{q})$ is **not** *LAS*.

Appendix C: Derivations for Proposition 2

These stationary equilibria $(p^*, q^*) = (1, \bar{q})$ and $(0, \tilde{q})$ are characterized by:

$$q^* = 1 + v(b) - v(\omega) - s(p^*, q^*), \tag{C.1}$$

and $u(b) \gtreqless s(p^*, q^*)$. Differentiating (C.1) with respect to the parameters such as b, λ, ω and γ, respectively, yields the following results. The effects of a change in the benefit level, b, in the *B&C* equilibrium $(1, \bar{q})$ and the *non-take-up* equilibrium $(0, \tilde{q})$ are respectively given by:

$$\frac{d\bar{q}}{db} = \frac{v'(b)}{1 + s_q(1, \bar{q})} > 0 \text{ and } \frac{d\tilde{q}}{db} = \frac{v'(b)}{1 + s_q(0, \tilde{q})} > 0, \tag{C.2}$$

while the population of deserving claimants, p, remains unaffected in either equilibrium because p is fixed at 0 or 1, where $1 + s_q(1, \bar{q}) > 0$ and $1 + s_q(0, \tilde{q}) > 0$. The effects of an increase in the degree of public exposure, λ, are given by:

$$\frac{d\bar{q}}{d\lambda} = -\frac{1}{1 + s_q(1, \bar{q})} \frac{\partial s(1, \bar{q})}{\partial \lambda} < 0 \text{ and } \frac{d\tilde{q}}{d\lambda} = -\frac{1}{1 + s_q(0, \tilde{q})} \frac{\partial s(0, \tilde{q})}{\partial \lambda} < 0, \tag{C.3}$$

where $\frac{\partial s(1,\bar{q})}{\partial \lambda} = \frac{1-\gamma}{2}\frac{\bar{q}(1-\bar{q})}{\gamma+(1-\gamma)\bar{q}} > 0$ and $\frac{\partial s(0,\tilde{q})}{\partial \lambda} = \frac{1-\tilde{q}}{2} > 0$. The effects of an increase in the wage rate ω in the respective stationary equilibria are given by:

$$\frac{d\bar{q}}{d\omega} = -\frac{v^{'}(\omega)}{1+s_q(1,\bar{q})} < 0 \text{ and } \frac{d\tilde{q}}{d\omega} = -\frac{v^{'}(\omega)}{1+s_q(0,\tilde{q})} < 0.$$

The effects of an increase in the population share of the deserving poor among the poor γ in the respective stationary equilibria are given by:

$$\frac{d\bar{q}}{d\gamma} = -\frac{1}{1+s_q(1,\bar{q})}\frac{\partial s(1,\bar{q})}{\partial \gamma} > 0 \text{ and } \frac{d\tilde{q}}{d\gamma} = -\frac{1}{1+s_q(0,\tilde{q})}\frac{\partial s(0,\tilde{q})}{\partial \gamma} < 0,$$

where $\frac{\partial s(1,\bar{q})}{\partial \gamma} = \frac{\lambda}{2}\frac{-1+\bar{q}[1-(1-\bar{q})[\gamma+(1-\gamma)\bar{q}]]}{[\gamma+(1-\gamma)\bar{q}]^2} < 0$ and $\frac{\partial s(0,\tilde{q})}{\partial \gamma} = \frac{\lambda}{2}\frac{1-(1-\tilde{q})(1-\gamma)\tilde{q}}{(1-\gamma)^2\tilde{q}} > 0$.

Appendix D: Proofs of Lemma 2

Proof of Claim (i). By totally differentiating, we obtain the slope of the locus $u(b) - s^r(p,q) = 0$:

$$\frac{dq}{dp} = -\frac{s_p^r(p,q)}{s_q^r(p,q)} < 0, \text{ with} \tag{D.1}$$

$$s_p^r(p,q) = \frac{\lambda^r}{2\beta\gamma\eta^2}\frac{1}{y-T} + \frac{\lambda^r}{2\gamma\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2}\frac{b}{1-\beta} > 0, \text{ and} \tag{D.2}$$

$$s_q^r(p,q) = \frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{1-\gamma}{y-T} + \frac{\lambda^r}{2\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2}\frac{b(1-\gamma)}{1-\beta} > 0. \tag{D.3}$$

Similarly, the slope of the locus $v(b) - v(\omega) + 1 - q - s^r(p,q) = 0$ is given by:

$$\frac{dq}{dp} = -\frac{s_p^r(p,q)}{1+s_q^r(p,q)}. \tag{D.4}$$

Comparing $(D.1)$ with $(D.4)$ reveals that $\frac{s_p^r(p,q)}{s_q^r(p,q)} > \frac{s_p^r(p,q)}{1+s_q^r(p,q)}$. ■

Proof of Claim (ii). We can rewrite the locus $v(b) - v(\omega) + 1 - q - s^r(p,q) = 0$, evaluated at $(1,\bar{q}^r)$, as $v(\omega) - v(b) = 1 - \bar{q}^r - s^r(1,\bar{q}^r)$. Since the left-hand side of the last expression is positive (recall $\omega > b$), the right-hand side must be positive, implying that $\bar{q}^r$ must be less than 1. ■

Proof of Claim (iii). Since setting $q = 1$ violates the equality of $v(\omega) - v(b) = 1 - \bar{q}^r - s^r(1,\bar{q}^r)$, q must be less than 1. ■

Appendix E: Proofs of Proposition 3

(i) The determinant of the Jacobian evaluated at $(\hat{p}^r, \hat{q}^r)$ is given by

$$\begin{aligned} & \begin{vmatrix} -\hat{p}^r(1-\hat{p}^r)s_p(\hat{p}^r,\hat{q}^r) & -\hat{p}^r(1-\hat{p}^r)s_q(\hat{p}^r,\hat{q}^r) \\ -\hat{q}^r(1-\hat{q}^r)s_p(\hat{p}^r,\hat{q}^r) & -\hat{q}^r(1-\hat{q}^r)\left[1+s_q(\hat{p}^r,\hat{q}^r)\right] \end{vmatrix} \\ = \; & \hat{p}^r(1-\hat{p}^r)\hat{q}^r(1-\hat{q}^r)s_p^r(\hat{p}^r,\hat{q}^r) > 0, \end{aligned}$$

while its trace is negative, because $s_p^r(\hat{p}^r,\hat{q}^r) > 0$ and $s_q^r(\hat{p}^r,\hat{q}^r) > 0$ (see $(D.2)$ and $(D.3)$ in Appendix D). Hence, $(\hat{p}^r,\hat{q}^r)$ is *LAS*.

(ii) The determinant evaluated at $(1,\bar{q}^r)$ is given by

$$\begin{aligned} & \begin{vmatrix} -[u(b)-s^r(1,\bar{q}^r)] & 0 \\ -\bar{q}^r(1-\bar{q}^r)s_p^r(1,\bar{q}^r) & -\bar{q}^r(1-\bar{q}^r)\left[1+s_q^r(1,\bar{q}^r)\right] \end{vmatrix} \\ & = [u(b)-s^r(1,\bar{q}^r)]\bar{q}^r(1-\bar{q}^r)\left[1+s_q^r(1,\bar{q}^r)\right] \gtreqless 0 \end{aligned}$$

with $s_q^r(1,\bar{q}^r) > 0$. Moreover if $u(b)-s^r(1,\bar{q}^r) > 0$, the determinant is positive and the trace is negative, thus implying that $(1,\bar{q}^r)$ is *LAS*. Conversely, if $u(b)-s^r(1,\bar{q}^r) < 0$, the determinant is negative and thus $(1,\bar{q}^r)$ is a saddle.

(iii) The determinant evaluated at $(0,\tilde{q}^r)$ is given by

$$\begin{aligned} & \begin{vmatrix} u(b)-s^r(0,\tilde{q}^r) & 0 \\ -\tilde{q}^r(1-\tilde{q}^r)s_p^r(0,\tilde{q}^r) & -\tilde{q}^r(1-\tilde{q}^r)\left[1+s_q^r(0,\tilde{q}^r)\right] \end{vmatrix} \\ = \; & -[u(b)-s^r(0,\tilde{q}^r)]\tilde{q}^r(1-\tilde{q}^r)\left[1+s_q^r(0,\tilde{q}^r)\right] \gtreqless 0 \end{aligned}$$

with $1+s_q^r(0,\tilde{q}^r) > 0$. Moreover, if $u(b)-s^r(0,\tilde{q}^r) < 0$, then the determinant is positive and the trace is negative, thus implying that $(0,\tilde{q}^r)$ is *LAS*. Conversely, if $u(b)-s^r(0,\tilde{q}^r) > 0$, then the determinant is negative, and thus $(0,\tilde{q}^r)$ is a saddle.

(iv) Since the determinant evaluated at $(\bar{p}_2^r,1)$ is given by

$$\begin{aligned} & \begin{vmatrix} -\bar{p}_2^r(1-\bar{p}_2^r)s_p^r(\bar{p}_2^r,1) & -\bar{p}_2^r(1-\bar{p}_2^r)s_q^r(\bar{p}_2^r,1) \\ 0 & -\left[v(b)-v(\omega)-s^r(\bar{p}_2^r,1)\right] \end{vmatrix} \\ = \; & \bar{p}_2^r(1-\bar{p}_2^r)s_p^r(\bar{p}_2^r,1)\left[v(b)-v(\omega)-s^r(\bar{p}_2^r,1)\right] < 0, \end{aligned}$$

$(\bar{p}_2^r,1)$ is a saddle.

(v) The determinant evaluated at $(\tilde{p}_2^r,0)$ is given by

$$\begin{aligned} & \begin{vmatrix} -\tilde{p}_2^r(1-\tilde{p}_2^r)s_p^r(\tilde{p}_2^r,0) & -\tilde{p}_2^r(1-\tilde{p}_2^r)s_q^r(\tilde{p}_2^r,0) \\ 0 & v(b)-v(\omega)+1-s^r(\tilde{p}_2^r,0) \end{vmatrix} \\ = \; & -\tilde{p}_2^r(1-\tilde{p}_2^r)s_p^r(\tilde{p}_2^r,0)\left[v(b)-v(\omega)+1-s^r(\tilde{p}_2^r,0)\right] \gtreqless 0 \end{aligned}$$

with $s_p^r(\tilde{p}_2^r, 0) > 0$. Moreover, if $v(b) - v(\omega) + 1 - s^r(\tilde{p}_2^r, 0) < 0$, the determinant is positive and the trace is negative, thus implying that $(\tilde{p}_2^r, 0)$ is *LAS*. Conversely, if $v(b) - v(\omega) + 1 - s^r(\tilde{p}_2^r, 0) > 0$, the determinant is negative and thus $(\tilde{p}_2^r, 0)$ is a saddle.

(vi) The determinant evaluated at $(1, 1)$ is given by

$$\begin{aligned} & \begin{vmatrix} -[u(b) - s^r(1,1)] & 0 \\ 0 & -[v(b) - v(\omega) - s^r(1,1)] \end{vmatrix} \\ = \; & [u(b) - s^r(1,1)][v(b) - v(\omega) - s^r(1,1)] \gtreqless 0. \end{aligned}$$

Since $v(b) - v(\omega) - s^r(1,1) < 0$, $(1, 1)$ is a saddle if $u(b) - s^r(1,1) > 0$, while it is a source if $u(b) - s^r(1,1) < 0$. On the other hand, the determinant evaluated at $(0, 0)$ is given by

$$\begin{vmatrix} u(b) - s^r(0,0) & 0 \\ 0 & v(b) - v(\omega) + 1 - s^r(0,0) \end{vmatrix} = u(b)\,[v(b) - v(\omega) + 1] \gtreqless 0,$$

due to $s^r(0,0) = 0$. Moreover, if $v(b) - v(\omega) + 1 > 0$, the determinant and the trace are both positive and thus $(0, 0)$ is a source. Conversely, if $v(b) - v(\omega) + 1 < 0$, it is a saddle. The determinant evaluated at $(0, 1)$ is given by

$$\begin{aligned} & \begin{vmatrix} u(b) - s^r(0,1) & 0 \\ 0 & -[v(b) - v(\omega) - s^r(0,1] \end{vmatrix} \\ = \; & -[u(b) - s^r(0,1)][v(b) - v(\omega) - s^r(0,1)] \gtreqless 0, \end{aligned}$$

whose trace is given by $u(b) - v(b) + v(\omega) > 0$, implying that $(0, 1)$ is either a saddle or a source.

(vii) The determinant evaluated at $(1, 0)$ is given by

$$\begin{aligned} & \begin{vmatrix} -[u(b) - s^r(1,0)] & 0 \\ 0 & v(b) - v(\omega) - s^r(1,0) \end{vmatrix} \\ = \; & -[u(b) - s^r(1,0)][v(b) - v(\omega) - s^r(1,0)] \gtreqless 0. \end{aligned}$$

Moreover, if $u(b) - s^r(1,0) > 0$ and $v(b) - v(\omega) + 1 - s^r(1,0) < 0$, the determinant is positive and the trace is negative; hence, $(1, 0)$ is *LAS*. Otherwise, it is either a source or a saddle.

(viii) Since the determinants evaluated at $(1, \tilde{q}_2^r)$, $(0, \tilde{q}_2^r)$ and $(\tilde{p}^r, 0)$ are all equal to zero, we cannot apply the linear approximation technique. Nevertheless, we can easily find that one of the eigenvalues is positive and thus all of them are unstable.

Appendix F: Derivations for Proposition 4

We first investigate the effect of changes in the benefit level, b, on the interior stationary point $(\hat{p}^r, \hat{q}^r)$ characterized by

$$u(b) - s^r(\hat{p}^r, \hat{q}^r) = 0, \text{ and} \tag{F.1}$$
$$v(b) - v(\omega) + 1 - \hat{q}^r - s^r(\hat{p}^r, \hat{q}^r) = 0. \tag{F.2}$$

The above two conditions are reduced to

$$v(b) - v(\omega) + 1 - \hat{q}^r - u(b) = 0. \tag{F.3}$$

Differentiating $(F.3)$ with respect to b yields

$$\frac{d\hat{q}^r}{db} = v^{'}(b) - u'(b) \gtreqless 0. \tag{F.4}$$

After differentiating $(F.1)$ with respect to b, substituting $(F.4)$ into the resultant expression yields

$$\frac{d\hat{p}^r}{db} = \frac{\left[1 + s_q^r(\hat{p}^r, \hat{q}^r)\right] u'(b) - s_q^r(\hat{p}^r, \hat{q}^r)v^{'}(b) - [\partial s^r(\hat{p}^r, \hat{q}^r)/\partial b]}{s_p^r(\hat{p}^r, \hat{q}^r)} \gtreqless 0,$$

where $\frac{\partial \hat{s}^r(\hat{p}^r,\hat{q}^r)}{\partial b} = \frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial T}\frac{\partial T(\hat{p}^r,\hat{q}^r)}{\partial} > 0$. Differentiating $(F.2)$ with respect to b, evaluated at $(1, \bar{q}^r)$, yields

$$\frac{d\bar{q}^r}{db} = \frac{v^{'}(b) - [\partial s^r(1, \bar{q}^r)/\partial b]}{1 + s_q^r(1, \bar{q}^r)} \gtreqless 0,$$

where $s_q^r(1, \bar{q}^r) > 0$ from $(D.3)$ and $\frac{\partial s^r(1,\bar{q}^r)}{\partial b} = \frac{\partial s^r(1,\bar{q}^r)}{\partial T}\frac{\partial T(1,\bar{q}^r)}{\partial b} > 0$ from (28). Differentiating $(F.2)$ with respect to b, evaluated at $(0, \tilde{q}^r)$, yields

$$\frac{d\tilde{q}^r}{db} = \frac{v^{'}(b) - [\partial s^r(0, \tilde{q}^r)/\partial b]}{1 + s_q^r(0, \tilde{q}^r)} \gtreqless 0,$$

where $s_q^r(0, \tilde{q}^r) > 0$ from $(D.3)$ and $\frac{\partial s^r(0,\tilde{q}^r)}{\partial b} = \frac{\partial s^r(0,\tilde{q}^r)}{\partial T}\frac{\partial T(0,\tilde{q}^r)}{\partial b} > 0$. Differentiating $(F.1)$ with respect to b, evaluated at $(\tilde{p}_2^r, 0)$, yields

$$\frac{d\tilde{p}_2^r}{db} = \frac{u^{'}(b) - [\partial s^r(\tilde{p}_2^r, 0)/\partial b]}{s_p^r(\tilde{p}_2^r, 0)} \gtreqless 0, \text{ if and only if } u^{'}(b) - [\partial s^r(\tilde{p}_2^r, 0)/\partial b] \gtreqless 0,$$

where $s_p^r(\tilde{p}_2^r, 0) > 0$ from $(D.2)$ and $\frac{\partial s^r(\tilde{p}_2^r,0)}{\partial b} > 0$.

The effects of changes in λ, evaluated at $(\hat{p}^r,\hat{q}^r)$, $(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, and $(\tilde{p}_2^r,0)$, are respectively given by

$$\frac{d\hat{p}^r}{d\lambda^r}=-\frac{1}{s_p^r(\hat{p}^r,\hat{q}^r)}\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial\lambda^r}<0,\ \frac{d\hat{q}^r}{d\lambda^r}=0,\ \frac{d\bar{q}^r}{d\lambda^r}=-\frac{1}{1+s_q^r(1,\bar{q}^r)}\frac{\partial s^r(1,\bar{q}^r)}{\partial\lambda^r}<0,$$
$$\frac{d\tilde{q}^r}{d\lambda^r}=-\frac{1}{1+s_q^r(0,\tilde{q}^r)}\frac{\partial s^r(0,\tilde{q}^r)}{\partial\lambda^r}<0,\text{ and }\frac{d\tilde{p}_2^r}{d\lambda^r}=-\frac{1}{s_q^r(\tilde{p}_2^r,0)}\frac{\partial s^r(\tilde{p}_2^r,0)}{\partial\lambda^r}<0, \tag{F.5}$$

since $\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial\lambda^r}=\frac{1}{2\beta\gamma^2\eta^2}\frac{\gamma\hat{p}^r+(1-\gamma)\hat{q}^r}{y-T}>0$ and $\frac{\partial s^r(p,q)}{\partial\lambda^r}>0$ for $(p,q)=(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, or $(\tilde{p}_2^r,0)$.

The effects of changes in the income level of the rich, y, evaluated at $(\hat{p}^r,\hat{q}^r)$, $(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, and $(\tilde{p}_2^r,0)$, are respectively given by

$$\frac{d\hat{p}^r}{dy}=-\frac{1}{s_p^r(\hat{p}^r,\hat{q}^r)}\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial y}>0,\ \frac{d\hat{q}^r}{dy}=0,\ \frac{d\bar{q}^r}{dy}=-\frac{1}{1+s_q^r(1,\bar{q}^r)}\frac{\partial s^r(1,\bar{q}^r)}{\partial y}>0,$$
$$\frac{d\tilde{q}^r}{dy}=-\frac{1}{1+s_q^r(0,\tilde{q}^r)}\frac{\partial s^r(0,\tilde{q}^r)}{\partial y}>0,\text{ and }\frac{d\tilde{p}_2^r}{dy}=-\frac{1}{s_p^r(\tilde{p}_2^r,0)}\frac{\partial s^r(\tilde{p}_2^r,0)}{\partial y}>0,$$

where $\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial y}=-\frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma\hat{p}^r+(1-\gamma)\hat{q}^r}{(y-T)^2}<0$ and $\frac{\partial s^r(p,q)}{\partial y}<0$ for $(p,q)=(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, or $(\tilde{p}_2^r,0)$.

The effects of changes in the wage rate, ω, evaluated at $(\hat{p}^r,\hat{q}^r)$, $(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, and $(\tilde{p}_2^r,0)$, are respectively given by

$$\frac{d\hat{p}^r}{d\omega}>0,\ \frac{d\hat{q}^r}{d\omega}=-v^{'}(\omega)<0,\ \frac{d\bar{q}^r}{d\omega}=-\frac{v^{'}(\omega)}{1+s_q^r(1,\bar{q}^r)}<0,$$
$$\frac{d\tilde{q}^r}{d\omega}=-\frac{v^{'}(\omega)}{1+s_q^r(0,\tilde{q}^r)}<0,\text{ and }\frac{d\tilde{p}_2^r}{d\omega}=0.$$

The effects of changes in the wage rate, η, evaluated at $(\hat{p}^r,\hat{q}^r)$, $(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, and $(\tilde{p}_2^r,0)$, are respectively given by:

$$\frac{d\hat{p}^r}{d\eta}=-\frac{1}{s_p^r(\hat{p}^r,\hat{q}^r)}\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial\eta}>0,\ \frac{d\hat{q}^r}{d\eta}=0,\ \frac{d\bar{q}^r}{d\eta}=-\frac{1}{1+s_q^r(1,\bar{q}^r)}\frac{\partial s^r(1,\bar{q}^r)}{\partial\eta}>0,$$
$$\frac{d\tilde{p}_2^r}{d\eta}=-\frac{1}{s_p^r(\tilde{p}_2^r,0)}\frac{\partial s^r(\tilde{p}_2^r,0)}{\partial\eta}>0,\text{ and }\frac{d\tilde{q}^r}{d\eta}=-\frac{1}{1+s_q^r(0,\tilde{q}^r)}\frac{\partial s^r(0,\tilde{q}^r)}{\partial\eta}>0,$$

where $\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial\eta}=-\frac{\lambda^r}{\beta\gamma^2\eta^3}\frac{\gamma\hat{p}^r+(1-\gamma)\hat{q}^r}{y-T}<0$ and $\frac{\partial s^r(p,q)}{\partial\eta}<0$ for $(p,q)=(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, or $(\tilde{p}_2^r,0)$.