CESifo WORKING PAPERS

8621
2020

October 2020

Replicator Evolution of Welfare Stigma: Welfare Fraud vs. Incomplete Take-Up

Jun-ichi Itaya, Kenichi Kurita

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8621*

Replicator Evolution of Welfare Stigma: Welfare Fraud vs. Incomplete Take-Up

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 using simple replicator models of statistical discrimination and the tax-payer resentment view welfare stigma suggested by Besley and Coate (1992). 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 model of statistical discrimination view welfare stigma, but it is stable in the model of the tax-payer resentment view welfare stigma. This difference arises from the different nature of stigma cost functions in these two models.

JEL-Codes: H310, H530, I380.

Keywords: stigma, replicator dynamics, incomplete take-up, welfare fraud, non-take-up.

*Jun-ichi Itaya**
Graduate School of Economics and Business Administration, Hokkaido University
Japan – Sapporo, 060-0809
itaya@econ.hokudai.ac.jp

Kenichi Kurita
Urban Institute and Department of Civil Engineering, School of Engineering ,
Kyushu University
744 Motooka, Nishi-ku
Japan – Fukuoka, 819-0395
kurita.kenichi.564@m.kyushu-u.ac.jp

*corresponding author

October 6, 2020
We are grateful to Toshihiro Ihori and Takuya Obara for helpful comments on earlier drafts. The usual disclaimers apply. The first author gratefully acknowledges JSPS KAKENHI Grant Number JP19K01711, while the second author gratefully acknowledges JSPS KAKENHI Grant Numbers JP19K23194 and JP20K13486.

1 Introduction

There are two important problems in welfare benefit programs: the prevalence of *welfare fraud* and *incomplete take-up*, whereby eligible poor people are reluctant to claim welfare benefits. This study investigates both problems in a simple framework of replicator dynamics with respect to heterogenous populations. In particular, we focus on how welfare stigma forms over time in terms of population dynamics. Although welfare stigma is important in understanding the emergence of welfare fraud and incomplete take-up, we find that the dynamic interaction between both eligible claimants and non-eligible claimants also matters in determining welfare stigma.

Table 1 shows take-up rates in Japan, the UK, the US, and Germany. The table reports that take-up rates are less than 1, indicating incomplete take-up.

	Data	Take up rate
Tachibanaki and Urakawa (2006)[JPN]	SIR	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).

Several empirical studies, on the other hand, find the existence of stigma in welfare programs. Moffitt (1983) examines the existence of welfare stigma in the Aid to Families with Dependent Children (AFDC) program using both a theoretical model and empirical analysis and finds significant welfare stigma in participants of the program. Bharagava and Manoli (2015) conduct a field experiment with the International Revenue Service Earned Income Tax Credit (EITC) program and investigate the sources of incomplete take-up. They conclude that welfare stigma has a statistically significant impact on the take-up rates of welfare benefits, although the effect is not so strong. Their findings are consistent with high take-up rate in the US, as seen in Table 1.

The existence of welfare stigma means that some people choose not to receive public assistance despite satisfying the eligibility criteria. As Table 1 indicates, eligible recipients make such a decision because they fear negative labels, disapproval, or public shaming if they participate in a public assistance program. Given the *negative* perception of claiming public assistance, it is not surprising that this lifeline was been viewed favorably, even when the social security system was established in Japan. Though the public assistance system is designed to complement the social insurance system, the low take-up rate of public assistance due to welfare stigma is problematic in that it creates inefficiency because it prevents the social security system from functioning as expected. This outcome 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 one of the major policy issues in welfare programs.

On theoretical grounds, Moffitt (1983), Besley and Coate (1992), 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), who allows for endogenous choices on whether or not to take up benefits, but not for welfare fraud. This study aims to fill this research gap by considering together *welfare fraud* and *incomplete take-up* in a simple replicator dynamics framework.

There are several innovations and findings in this study. First, we introduce endogenous choices of take-up of welfare recipients into Besley and Coate's (1992) 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 (i.e., low take-up rates) emerges endogenously, which Besley and Coate (1992) 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 replicator dynamics model that endogenously and jointly determines the populations of *welfare fraud* and *incomplete take-up* through time. Second, Besley and Coate (1992) 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 expectations equilibria in the sense that all individuals can precisely predict the stigma costs in equilibrium. Consequently, no further revisions to know the precise amount of stigma costs take place at that equilibrium in a sort of thought experiment. This scenario requires a great deal of common knowledge and god-like calculation powers in a timeless world. In addition, a compelling reason or satisfactory motivation for welfare claimants stick to finding a *stationary value* (or fixed point) of stigma costs seems unclear to us. Specifically, the explanatory significance of equilibrium concepts in general depends on the *plausibility* of the underlying dynamics that bring the players to equilibrium. The central role of this equilibrium concept in Besley and Coate's (1992) analysis, that is, their timeless scenario on how to reach such an equilibrium is not especially convincing to describe the actual behavior of welfare beneficiaries.

In contrast, applying replicator dynamics allows us to explain how the level of stigma costs is formed through the population dynamics in terms of the deserving and undeserving claimants instead of the process of thought experiments in terms of stigma costs themselves. Moreover, the replicator dynamics model is more compatible with the model of the *statistical discrimination view stigma*, although it is intrinsically a static one. This is because the size of the stigma cost is evidently sensitive to the population profile of the deserving and undeserving claimants in the sense that higher populations of the current welfare fraud induces more fraud, thereby increasing stigma costs.

Third, our replicator dynamics give rise to multiple long-run equilibria, which Besley and Coate (1992) do not address. The multiplicity of long-run equilibria are caused by aggregate externalities arising from changes in the *composition* of the heterogenous populations of the poor through 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 statics may emerge depending on the resulting stationary equilibria. Thus, the comparative statics analysis carried out by Besley and Coate (1992) may lead to misleading predictions because they focus only on a unique fixed point.

Fourth, and closely related to third point, the replicator dynamics provide a method to refine the equilibrium, even if multiple equilibria exist. The stability analysis of the replicator 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, that is, 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 take up welfare benefits) and the other that allows for welfare fraud to some extent and 100% incomplete take-up (i.e., no deserving poor individuals take up welfare benefits). A comparative statics analysis in either boundary stationary equilibrium reveals that although the population of deserving claimants is unaffected by any parameter changes in the long run, that of undeserving claimants rises in response to increasing welfare benefits, as well as reductions in the degree of public exposure and wage rates. The most important policy implication is that since the first equilibrium with full take-up is more socially desirable compared to the second equilibrium with 100% incomplete take-up, increasing welfare benefits makes it more likely that the society will reach the first better equilibrium autonomously in the long run.

Fifth, we construct a replicator dynamic model of *tax-payer resentment view stigma.* Although there also exist multiple long-run equilibria, we can use the stability analysis to pin down a **unique** stable long-run equilibrium; that is, an interior stationary point or either of the above-mentioned boundary stationary equilibria. The most important difference is that in the tax-payer resentment model, an interior stationary equilibrium is stable, which implies that it is more likely to allow for the **coexistence** of welfare fraud and incomplete take-up.

The organization of the paper is as follows. The next section and Section 3 describe the basic model and then characterize the replicator dynamics for the *statistical discrimination view stigma* model. Sections 4 and 5 conduct a stability analysis and a comparative static analysis with respect to the principal parameters, respectively, to address policy implications. Section 6 performs the same analysis in the replicator dynamic model of the tax-payer resentment view stigma. Section 5 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

We consider a society composed of two income classes: the poor and the rich. We normalize the total population to be equal to 1. The population of the poor income class is $\beta \in (0, 1)$, while the proportion of the rich income class is $1 - \beta$. People in the rich income class have incomes of y. The poor income class is further divided into two types: the deserving (*needy*) poor and undeserving (*non-needy*) 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. The populations of the deserving and undeserving poor are $\beta\gamma$ and $\beta(1-\gamma)$, respectively. For analytical simplicity, we assume that both β and γ are constant through time.

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 both the deserving and undeserving poor suffer from *stigma costs* when they receive welfare benefits. Taken together, the payoffs to the deserving and undeserving poor are

		Undeserving poor	
		Take-up	Work
Deserving Poor	Take-up	$u(b) - s(p,q), v(b) - s(p,q)$	$u(b) - s(p,q), v(\omega) - \theta$
	Non-take-up	$0, v(b) - s(p,q)$	$0, v(\omega) - \theta$

Table 2 Payoffs

where $s(p, q)$ represents the stigma costs suffered when taking up benefits. In addition, $u(\cdot)$ and $v(\cdot)$ represent the utility functions of the deserving and undeserving poor in terms of their own consumption, respectively, both of which are of C^2 class, strictly increasing in their own consumption and concave. Without loss of generality, we assume $y > \omega > b > 0$, where the second inequality guarantees that the undeserving poor are willing to work rather than taking up benefits if the disutility arising from labor supply were to be relatively low.

As in Besley and Coate (1992), we also assume that the degree of the disutility of work among the undeserving poor is monotonically increasing in θ over the interval $[0, 1]$ (see Fig. 2 also). Then, we define a *threshold* value of θ for the undeserving poor, $\hat{\theta}$, in the decision to work or not, as follows:

$$v(\omega) - \hat{\theta} = v(b) - s.$$

All individuals in the undeserving poor group who have $\theta \geq \hat{\theta}$ prefer to take up welfare benefits over working, and vice versa. Thus, the population share of undeserving poor who *want* to take up benefits is

$$\Pr(\text{Take-up benefit}|\text{Undeserving poor}) = \Pr(\theta \geq \hat{\theta}) = 1 - \hat{\theta}.$$

Note that there is an important difference between Besley and Coate's model and the present model in that the population of *actual* or *current* undeserving poor welfare claimants (who we may call *welfare fraud*), denoted by q, is always equal to $1 - \hat{\theta}$ in their static model, whereas in the present dynamic model, q evolves over time and is generally different from $1 - \hat{\theta}$ in every moment in time (see Fig. 2 also).

Next, we formulate a stigma cost function. We first employ the *statistical discrimination view stigma* suggested by Besley and Coate (1992), which is an increasing function of the discrepancy between the average disutility of work among all welfare claimants, denoted by $\bar{\theta}_w$, and the average disutility of work among the poor, denoted by $\bar{\theta}$, who consist of deserving and undeserving claimants. To obtain closed-form solutions for q and p, we specify the stigma cost function as a *linear* function of the difference $\bar{\theta}_w - \bar{\theta}$:

$$s(\bar{\theta}_w - \bar{\theta}) = \lambda[\bar{\theta}_w - \bar{\theta}], \tag{1}$$

where the constant parameter λ measures the degree of public exposure in welfare programs (i.e., when $\lambda = 0$ in (1), the welfare program is discretionary, while when $0 < \lambda < 1$, public exposure is *partial* when claiming welfare benefits and entails stigma costs), and the average disutility of work among the poor $\bar{\theta}$ is $\bar{\theta} = \int_0^1 \theta d\theta = 1/2$. The average disutility $\bar{\theta}_w$ among all welfare claimants is therefore determined according to

$$\bar{\theta}_w = \pi\bar{\theta}_d + (1-\pi)\bar{\theta}_u, \tag{2}$$

where

$$\pi := \Pr[\text{Deserving}|\text{Taking-up benefit}] = \frac{\gamma p}{\gamma p + (1-\gamma)q}, \tag{3}$$

$$\bar{\theta}_d := E[\theta|\text{Deserving} \cap \text{Taking-up benefit}], \text{ and}$$

$$\begin{aligned} \bar{\theta}_u \quad : \quad &= E[\theta|\text{Undeserving} \cap \text{Taking-up benefit}], \\ &= \int_{1-q}^{1} \frac{\theta d\theta}{q} = \left[\frac{\theta^2}{2q}\right]_{1-q}^{1} = 1 - \frac{q}{2}, \end{aligned} \tag{4}$$

where $\bar{\theta}_d$ represents the average disutility of work among the deserving poor and $\bar{\theta}_u$ is the average disutility of work among the undeserving claimants, while p and q represent the populations of *actual* deserving and undeserving poor claimants, respectively, both of which evolve over time as state variables of the replicator dynamics described later.[1] For simplicity, we assume not only that $\bar{\theta}_u > \bar{\theta}_d$ (i.e., the disutility of work among the undeserving claimants is greater than that among the deserving claimants), but also that the average disutility among the rich is the same as that of the poor, as in Besley and Coate (1992); consequently, we can set $\bar{\theta}_d = \bar{\theta} = 1/2$. Fig. 1 shows that π in (2) corresponds to the ratio between the upper shaded area (the deserving claimants) relative to the total shaded area (all welfare claimants).

Using (2), (3) and (4), together with $\bar{\theta}_d = \bar{\theta} = 1/2$, we can express the difference in (1) by

$$\bar{\theta}_w - \bar{\theta} = (1-\pi)\left(\bar{\theta}_u - \frac{1}{2}\right) = \frac{1}{2}\frac{(1-\gamma)q(1-q)}{\gamma p + (1-\gamma)q}. \tag{5}$$

[1] The variables p and q represent the **share** of the deserving poor who receive benefits relative to the total poor population (i.e., $(1-\beta)\gamma$) and the **share** of the undeserving poor who receive benefits relative to the total poor population (i.e., $(1-\beta)(1-\gamma)$), respectively. To avoid an abuse of language, we simply refer to p and q as the populations of the deserving and undeserving claimants, respectively, in what follows.

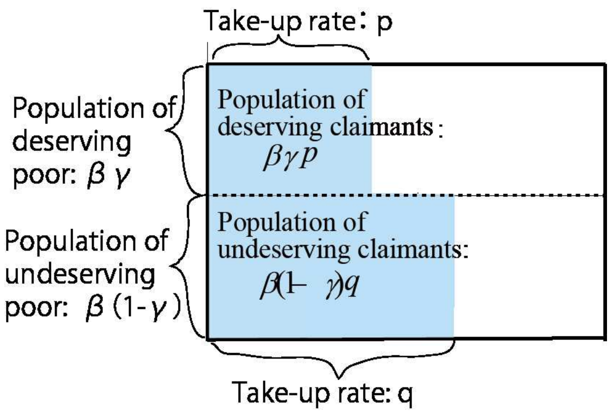

Figure 1: The shaded area represents the population of all welfare claimants

By substituting (5) into (1), we can rewrite the stigma cost function (1) as follows:

$$s(p, q) = \frac{\lambda}{2} \frac{(1-\gamma)q(1-q)}{\gamma p + (1-\gamma)q}. \tag{6}$$

As (6) makes clear, the stigma cost is affected not only by the parameters b, ω, λ, and γ, but also by the populations of deserving and undeserving claimants. Consequently, although the state variables p and q are fixed at each moment in time, they evolve over time, so the stigma cost level itself is also changing over time. Note, however, that the point $(p, q) = (0, 0)$ should not be contained in the domain of (p, q) because $s(p, q)$ is discontinuous at $(0, 0)$.[2]

3 Replicator dynamics

To capture the population dynamics of the state variables p and q, we construct the replicator dynamics of p and q as follows:

[2] It is easy to show this. Consider, for example, a case where q converges to 0 following $q = ap$ with a being any positive real number. Take the limit of (6):

$$\lim_{(p,q)\to(0,\ 0)} s(p, q) = \lim_{(p,q)\to(0,\ 0)} \frac{\lambda}{2} \frac{(1-\gamma)(ap)(1-q)}{\gamma p + (1-\gamma)(ap)} = \frac{\lambda}{2} \frac{(1-\gamma)a(1-q)}{\gamma + (1-\gamma)a}.$$

Depending on the arbitrarily chosen values of a, the limiting value of $s(p, q)$ takes different values.

$$\begin{aligned}\dot{p} &= p\left\{E[U_{\text{deserving poor}} \mid \text{Taking-up benefit}] - E[U_{\text{deserving poor}}]\right\}, \qquad (7)\\ &= p(1-p)[u(b) - s(p,q)],\end{aligned}$$

where

$$\begin{aligned}E[U_{\text{deserving poor}} \quad &\mid \quad \text{Taking-up benefit}] = q\left[u(b) - s(p,q)\right] + (1-q)\left[u(b) - s(p,q)\right],\\ &= \quad u(b) - s(p,q), \text{ and}\end{aligned}$$

$$\begin{aligned}E[U_{\text{deserving poor}}] &= pq\left[u(b) - s(p,q)\right] + p(1-q)\left[u(b) - s(p,q)\right]\\ &\quad +(1-p)q\cdot 0 + (1-p)(1-q)\cdot 0,\\ &= p\left[u(b) - s(p,q)\right].\end{aligned}$$

Next, the population growth rate of q is

$$\begin{aligned}\dot{q} &= q\left\{E[U_{\text{undeserving poor}} \mid \text{Taking-up benefit}] - E[U_{\text{undeserving poor}}]\right\},\\ &= q(1-q)\left[v(b) - s(p,q) - v(\omega) + \theta\right], \qquad (8)\end{aligned}$$

where

$$\begin{gathered}E[U_{\text{undeserving poor}} \mid \text{Taking-up benefit}]\\ = p\left[v(b) - s(p,q)\right] + (1-p)\left[v(b) - s(p,q)\right] = v(b) - s(p,q), \text{ and}\end{gathered}$$

$$\begin{aligned}E[U_{\text{undeserving poor}}] &= pq\left[v(b) - s(p,q))\right] + (1-p)q\left[v(b) - s(p,q)\right]\\ &\quad +p(1-q)\left[v(\omega) - \theta\right] + (1-p)(1-q)\left[v(\omega) - \theta\right],\\ &= q\left[v(b) - s(p,q)\right] + (1-q)\left[v(\omega) - \theta\right].\end{aligned}$$

Since θ varies across the undeserving poor; that is, it is uniformly distributed over the closed interval $[0, 1]$, we must choose an appropriate value for θ to describe the replicator dynamics of the model. Given q at each instant in time, there exists an undeserving poor individual who has $\theta = 1 - q$ (whom we call a *pivotal individual*). Moreover, if q is in the position depicted in Fig. 2, then $v(b) - s(p,q) > v(w) - (1-q)$, giving rise to $\dot{q} > 0$ due to (8). Consequently, the population of undeserving claimants q is rising in time. In particular, if $\hat{\theta} = 1 - q$ holds, then $\dot{q} = 0$ in (8). In economic terms, the pivotal individual is no longer replaced by other *new* individuals with further lower disutility to work so that q ceases to change. Hence, we could view the resulting long-run equilibrium in the present dynamic model as the static equilibrium analyzed by Besley and Coate (1992). These considerations lead

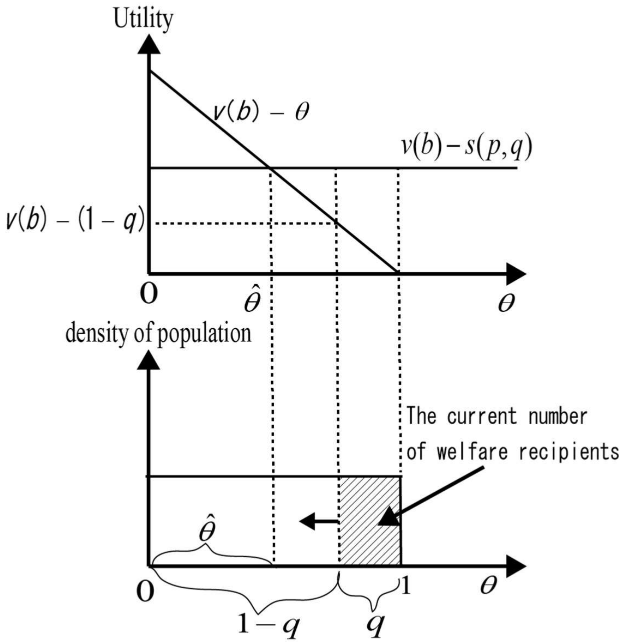

Figure 2: Pivotal person with $\theta = 1 - q$

us to set θ equal to $1 - q$. We again emphasize that such a *pivotal individual* with $\theta = 1-q$ is continuously replaced during the adjustment process towards the long-run equilibrium unless $\hat{\theta} = 1 - q$ holds.

We are now ready to analyze the behavior of the replicator dynamics in terms of p and q using a phase diagram method. To this end, we draw the graphs for the loci of points by setting $\dot{p} = 0$ in (7) and $\dot{q} = 0$ in (8) within the unit square $[0, 1] \times [0, 1]$ of R^2_+. That is,

$$p^* = 0, \; p^* = 1, \; u(b) - s(p^*, q^*) = 0, \tag{9}$$

$$q^* = 0, \; q^* = 1, \; v(b) - v(\omega) + 1 - q^* - s(p^*, q^*) = 0, \tag{10}$$

where p^* and q^* denote the stationary values of the state variables p and q, respectively. For the reason stated above, we exclude the stationary point $(0, 0)$.

To simplify the analysis, we make the following assumptions:

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

Assumption 2 $\lambda < 2$.

An immediate consequence of Assumptions 1 and 2 is $u(b) < 1$.

To depict the dynamic behavior of p and q, we first solve $u(b) - s(p,q) = 0$ (i.e., the locus $\dot{p} = 0$) for p, together with (6), to obtain

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

whose right-hand side is a quadratic function of q. The locus of this quadratic function is illustrated by the locus $\dot{p} = 0$ in the unit square of the (p,q) space of Figs. 3-5. This graph crosses the origin $(p,q) = (0,0)$ and the q-axis at $\tilde{q}_2 = -(2u(b)/\lambda) + 1 \in (0,1)$. Next, we solve $v(b) - v(\omega) + 1 - q - s(p,q) = 0$ (i.e., the locus $\dot{q} = 0$) for p, together with (6), to obtain

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

whose right-hand side is a rational function of q. This rational function is illustrated by the locus $\dot{q} = 0$ in the (p,q) space of Figs. 3-5. This graph crosses the origin $(p,q) = (0,0)$ and the q-axis at

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

Note that the intercept $\tilde{q}$ may or may not be greater than that of $\tilde{q}_2$.

Taken together, we can draw Figs. 3-5. Although varying the values of parameters such as λ, γ, b, and ω will yield different pictures for the loci of $\dot{p} = 0$ and $\dot{q} = 0$, we can depict **three** typical phase portraits capturing all possible qualitative movements of p and q. We should note that the loci $u(b) - s(p,q) = 0$ and $v(b) - v(\omega) + 1 - q - s(p,q) = 0$ may or may not intersect within the unit square of the space (p,q)

4 Stability

In this section, we investigate the stability properties of the stationary points of the system (7) and (8). Although there are many stationary points, we must focus on only *stable* stationary points to perform a meaningful comparative statics analysis.

For the stability analysis, we take a linear approximation of (7) and (8) around the stationary point (p^*, q^*):

$$\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} \partial f(p^*, q^*)/\partial p & \partial f(p^*, q^*)/\partial q \\ \partial g(p^*, q^*)/\partial p & \partial g(p^*, q^*)/\partial q \end{bmatrix} \begin{bmatrix} p^* - p \\ q^* - q \end{bmatrix}, \tag{11}$$

where the functions $f(\cdot)$ and $g(\cdot)$, respectively, represent the right-hand sides of (7) and (8), and

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

$$\frac{\partial f(p^*, q^*)}{\partial q} = -p^*(1-p^*)s_q(p^*, q^*),$$

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

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

noting that the stationary equilibrium values of p^* and q^* are given by any combination of (9) and (10) except $(0, 0)$.

Differentiating (6) with respect to p and q, respectively, yields

$$s_p(p,q) \equiv \frac{\partial s(p,q)}{\partial p} = -\frac{\lambda(1-\gamma)}{2}\frac{\gamma q(1-q)}{\left[\gamma p + (1-\gamma)q\right]^2} < 0, \text{ and} \tag{12}$$

$$s_q(p,q) \equiv \frac{\partial s(p,q)}{\partial q} = \frac{\lambda(1-\gamma)}{2}\frac{\gamma p(1-2q) - (1-\gamma)q^2}{\left[\gamma p + (1-\gamma)q\right]^2} \gtreqless 0. \tag{13}$$

Figs. 3-5 provide the phase portraits for this system depending on the relative locations of loci $\dot{p} = 0$ and $\dot{q} = 0$.

On the stability properties of the linearized system (11) around the respective stationary points listed below, we can demonstrate the following proposition:

Proposition 1 *Under the replicator dynamics* (7) *and* (8) *coupled with Assumptions 1 and 2 for all* $(p, q) \in [0, 1]^2 \setminus (0, 0)$*:*

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

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

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

(iv) The stationary point $(1, 0)$ *is locally asymptotically stable if* $v(b) - v(w) + 1 < 0$*. Conversely, if* $v(b) - v(w) + 1 > 0$*, then it is a saddle.*

(v) The stationary point $(1, 1)$ *is a saddle and* $(0, 1)$ *is a source.*

(vi) The stationary point $(0, \tilde{q}_2)$ *is **not** locally asymptotically stable.*[3]

[3] The stationary point $(0, \tilde{q}_2)$ could be a saddle; however, even if it is, it is unstable.

Proof. See Appendix A.[4] ■

Notice first that if a stationary point is a saddle, then the trajectories besides the ones starting from an initial condition lying on the one-dimensional stable manifold can reach that stationary point; thus it is reasonable to conclude that it is unstable. Hence, a stationary point that is either a saddle or a source is unstable. We first consider Fig. 3, in which there is **only one** stable stationary equilibrium point $(1, \bar{q})$. Indeed, since this stationary point is located above the locus $\dot{p} = 0$; that is, $\dot{p} > 0$, it implies that $u(b) - s(1, \bar{q}) > 0$ from (7). According to Claim (i) of Proposition 1, the stationary point $(1, \bar{q})$ is locally asymptotically stable. In contrast, since $(0, \tilde{q})$ in Fig. 3 is also located above the locus $\dot{p} = 0$, $\dot{p} > 0$. Since this implies that $u(b) - s(0, \tilde{q}) > 0$ in (7), $(0, \tilde{q})$ is unstable according to Claim (ii) of Proposition 1. All other stationary points in Fig. 3 are unstable due to Proposition 1.

As Fig. 4 illustrates, on the other hand, there coexist **two** stable stationary points, such as $(1, \bar{q})$ and $(0, \tilde{q})$. Both stationary points are locally asymptotically stable due to Claims (i) and (ii), respectively.

In Fig. 5, the locations of $\dot{p} = 0$ and $\dot{q} = 0$ are in reversed positions relative to Fig. 3. As Fig. 5 shows, there is only one stable stationary point $(0, \tilde{q})$ due to Claim (ii). According to Claim (iv), the stationary point $(1, 0)$ could be locally asymptotically stable. Nevertheless, Figs. 3-5 demonstrate that $(1, 0)$ is always unstable, because $v(b) - v(\omega) + 1 > 0$ holds in all figures.

Further, three remarks are in order. First, if there exists only one *stable* stationary equilibrium, it is unique and thus *globally* asymptotically stable. Hence, the system reaches a unique long-run equilibrium, such as $(1, \bar{q})$ or $(0, \tilde{q})$, independently of the initial condition. We call the former one the *Besley and Coate equilibrium* (or simply, the *B&C* equilibrium) and the latter one the *non-take-up equilibrium*.

Second, as all figures show, there are multiple stationary equilibria. Nevertheless, the stability analysis can help reduce the set of multiple stationary equilibria substantially, enabling us to pin down fewer stationary points: that is, **one** or **two** stable stationary equilibria, as we see in Figs. 3-5. In particular, two stable stationary equilibria emerge in Fig. 4. The emergence of multiple *stable* stationary equilibria indicates that, depending on the exogenously given initial values of p and q, the trajectories starting from different initial conditions could converge to different stationary points in the long run. This feature entails different comparative statics properties. In other

[4] 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.

words, we should not expect a unique prediction concerning the long-run comparative statics effects with respect to the structural parameters of the model or the policy instruments, unlike in Besley and Coate (1992). 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, the stigma cost is relatively lower when p is higher (recalling (6)), thereby inducing more of the deserving poor to take up welfare benefits. Thus, the population of deserving claimants continues to rise through time and therefore *all* deserving poor ultimately end up taking up welfare benefits in the long run.

Conversely, if the initial population of p is relatively small, then the stigma cost is higher due to $\partial s(0, \tilde{q})/\partial p < 0$ (recalling (12)), so that the process above is reversed. Consequently, no deserving claimants ultimately take up benefits in the long run (i.e., $(0, \tilde{q})$), which Besley and Coate (1992) do not examine in their study. Such a knife-edge property stems from the above-mentioned self-enforcing feature of the statistical discrimination model, which ultimately leads to one of the extreme 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 emerges from the monotonically decreasing stigma cost function of the population of deserving claimants, p, recalling (12). Therefore, initially larger values of p are associated with lower stigma costs, which further raise p, thus reducing stigma costs, and so on. In this way, this divergent process continues until p reaches the upper limit 1.[5] Conversely, since the effect of changes in q on the stigma cost is ambiguous in general (recall (13)), the relation between the population of undeserving claimants q and the stigma cost is **not monotonic**, and thus the self-enforcing process in terms of q does not arise. Thus, the long-run population of undeserving claimants tends to display intermediate values rather than either 0 or 1.

Third, our stability analysis reveals that the *interior* stationary point, which allows for the coexistence of welfare fraud and incomplete take-up, is unstable as a result of the property of cumulative divergence towards either of the extreme boundary stationary equilibria such as $(0, \tilde{q})$ or $(1, \bar{q})$. This striking and somewhat surprising result indicates that the coexistence of welfare fraud and incomplete take-up **never** emerges in the long run of the statistical discrimination model. 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] If the stigma cost function in (1) is concave in $\bar{\theta}_w - \bar{\theta}$, then this process becomes milder, while if the stigma cost function in (1) is convex in $\bar{\theta}_w - \bar{\theta}$, then this process will accelerate. In any case, this divergent property remains valid.

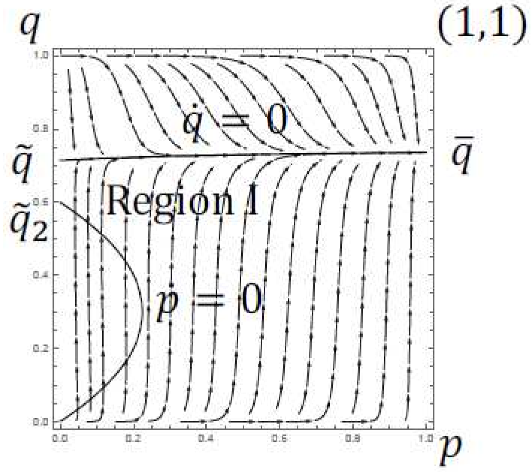

Figure 3: $\lambda = 0.25$, $u(b) = 0.05$, $v(b) = 0.25$, $v(\omega) = 0.5$, and $\gamma = 0.5$.

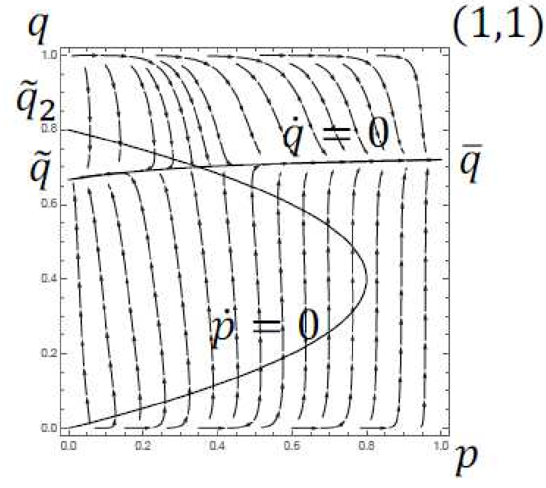

Figure 4: $\lambda = 0.5$, $u(b) = 0.05$, $v(b) = 0.25$, $v(\omega) = 0.5$, and $\gamma = 0.5$.

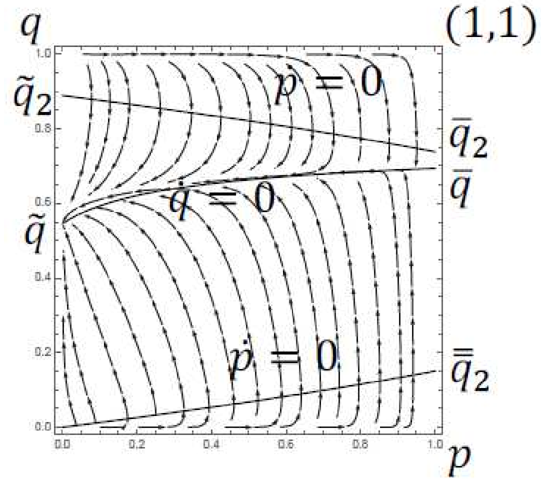

Figure 5: $\lambda = 0.9$, $u(b) = 0.05$, $v(b) = 0.25$, $v(\omega) = 0.5$, and $\gamma = 0.5$.

5 Comparative Statics

Based on the stability results obtained in the previous section, we confine our attention to only the *stable* long-run equilibria, such as $(0, \tilde{q})$ or $(1, \bar{q})$, to make a meaningful comparative static analysis. These stationary equilibria (together with $\theta = 1 - q$) are characterized by:

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

and $u(b) \gtreqless s(p^*, q^*)$, where $(p^*, q^*) = (1, \bar{q})$ or $(0, \tilde{q})$.

We first consider the effect of a change in the benefit level, b, on the *B&C* equilibrium $(1, \bar{q})$. Since p is fixed at 1, a change in b can affect **only** the population of undeserving claimants, q, in the long run. Taking this condition into account, we totally differentiate (14) with respect to b to obtain

$$\frac{d\bar{q}}{db} = \frac{v'(b)}{1 + s_q(1, \bar{q})} > 0,$$

while in the *non-take-up* equilibrium, $(0, \tilde{q})$, the effect is

$$\frac{d\tilde{q}}{db} = \frac{v'(b)}{1 + s_q(0, \tilde{q})} > 0,$$

noting that $s\,(p, q)$ does not depend directly on b. Moreover, from (A.4) and (A.6) in Appendix A, it follows that $1 + s_q(1, \bar{q}) > 0$ and $1 + s_q(0, \tilde{q}) > 0$.

Setting $p = 1$, the effect of an increase in the degree of public exposure, λ, on $\bar{q}$ in the *B&C* equilibrium, $(1, \bar{q})$, is

$$\frac{d\bar{q}}{d\lambda} = -\frac{1}{1 + s_q(1, \bar{q})} \frac{\partial s(1, \bar{q})}{\partial \lambda} < 0,$$

since $\frac{\partial s(1,\bar{q})}{\partial \lambda} = \frac{1-\gamma}{2} \frac{\bar{q}(1-\bar{q})}{\gamma+(1-\gamma)\bar{q}} > 0$, while in the *non-take-up* equilibrium, $(0, \tilde{q})$, is the effect of

$$\frac{d\tilde{q}}{d\lambda} = -\frac{1}{1 + s_q(0, \tilde{q})} \frac{\partial s(0, \tilde{q})}{\partial \lambda} < 0,$$

since $\frac{\partial s(0,\tilde{q})}{\partial \lambda} = \frac{1-\tilde{q}}{2} > 0$.

Next, we investigate the effect of increasing the wage rate, ω. In the *B&C* equilibrium, $(1, \bar{q})$, the effect on $\bar{q}$ is

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

while in the *non-take-up* equilibrium, $(0, \tilde{q})$, the effect is

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

noting that the function $s(p, q)$ does not depend directly on ω.

Finally, we consider the effect of an increase in the population share of the deserving poor among the poor, γ. In the *B&C* equilibrium, $(1, \bar{q})$, the effect on $\bar{q}$ is

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

since $\frac{\partial s(1,\bar{q})}{\partial \gamma} = \frac{\lambda}{2} \frac{-1+\bar{q}[1-(1-\bar{q})[\gamma+(1-\gamma)\bar{q}]]}{[\gamma+(1-\gamma)\bar{q}]^2} < 0$, while in the *non-take-up* equilibrium, $(0, \tilde{q})$, it is

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

since $\frac{\partial s(0,\tilde{q})}{\partial \gamma} = \frac{\lambda}{2} \frac{1-(1-q)(1-\gamma)q}{(1-\gamma)^2 q} > 0$.

To sum up,

Proposition 2 *Under Assumptions 1 and 2:*

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

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

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

(iv) *An increase in the proportion of deserving poor relative to the entire poor population increases the population of undeserving claimants in the B&C equilibrium, while it reduces that in the non-take-up equilibria.*

(v) *No parameter change affects the population of deserving claimants in 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, $b + \Delta b$, induces more of both the deserving and undeserving poor to take up the benefits because their utilities of receiving the benefits (i.e., $v(b)$ and $u(b)$) are higher than it would be otherwise. Although the population of undeserving claimants q is certainly rising according to (7) over time, that of deserving claimants p is **not**. The reason for the latter result is explained as follows. Since the *stable* stationary point $(0, \tilde{q})$ is initially located inside the area where $\dot{p} = p(1-p)\left[u(b) - s(0, \tilde{q})\right] < 0$ in Figs. 4 and 5, $\dot{p} < 0$ continues to hold after the increase in b. In other words, although the increased $u(b)$ would mitigate the downward pressure on $\dot{p}$, $\dot{p} < 0$ still holds as long as the increment of b is arbitrarily small. However, p must stay at 0 in a new stationary point because p initially reached the lower limit of $p = 0$, and thus it is impossible to fall further. More intuitively, since the increased b lowers $\hat{\theta} = v(\omega) - v(b + \Delta b) + s(0, \tilde{q} + \Delta q)$ in the new stationary point $(0, \tilde{q} + \Delta q)$ and since $s(0, \tilde{q} + \Delta q) < s(0, \tilde{q})$ due to $\partial s\left(0, \tilde{q}\right)/\partial q < 0$ in (13), more of the undeserving claimants with *lower disutility of work* are willing to take up benefits rather than work, so the average disutility of undeserving claimants θ_u falls (recalling (4)). Accordingly, $\bar{\theta}_w = \pi\bar{\theta}_d + (1-\pi)\bar{\theta}_u = \theta_u$ unambiguously falls because $p = 0$, implying that $\pi = 0$. Thus, the discrepancy $\bar{\theta}_w - \bar{\theta}$ (i.e., $\bar{\theta}_w - (1/2)$) unambiguously falls, as does the stigma cost. This is consistent with the resulting higher value of q.

Next, consider the stable stationary point $(1, \bar{q})$. As before, the increase in b induces more of both the deserving and undeserving poor to take up benefits. However, although q is certainly rising according to (8), p is **not** increasing. Since the stable stationary point $(1, \bar{q})$ is initially located outside the area where $\dot{p} = p(1-p)\left[u(b) - s(1, \bar{q})\right] > 0$ in Figs. 3 and 4, $\dot{p} > 0$ continues to hold after the small increase in b. Nevertheless, p must stay at 1 in the new stationary point $(1, \bar{q} + \Delta q)$ because it is impossible for it to rise further. More intuitively, the increase in b lowers $\hat{\theta} = v(\omega) - v(b + \Delta b) + s(1, \bar{q} + \Delta q)$ in the new stationary point in spite of $s(1, \bar{q} + \Delta q) \gtreqless s(1, \bar{q})$ due to (13) because the direct effect of increasing b on $\hat{\theta}$ dominates the indirect effect of the induced change in q through varying the stigma cost $s(1, \bar{q}+\Delta q)$. Consequently, q rises in the new stationary equilibrium.

Consider an increase of λ in the stable *non-take-np* equilibrium $(0, \tilde{q})$. An increase in λ directly raises the stigma cost (recall (6)), which discourages the incentive for the deserving and undeserving poor to take up benefits. According to (7) and (8), therefore, the populations of deserving claimants p and undeserving claimants q both tend to fall in time. Although q certainly falls, p does **not** decrease. Since the stable stationary point $(0, \tilde{q})$ is initially located inside the area where $\dot{p} = p(1-p)\left[u(b) - s(0, \tilde{q})\right] < 0$ in Figs. 4 and 5, $\dot{p} < 0$ is further strengthened by the increase in λ and the increased stigma cost ($\partial s(0, \tilde{q})/\partial q < 0$ due to (13)), but it is impossible for q to fall further. More intuitively, $\hat{\theta} = v(\omega) - v(b) + s(0, \tilde{q} - \Delta q)$ falls due to the increased stigma cost. The resulting decreased $\hat{\theta}$ is consistent with higher q in the new stationary equilibrium. An increase in the wage rate also directly reduces q according to (8) because higher wage rates induce more of the undeserving poor to work. The resulting decrease in q raises the stigma cost around $(0, \tilde{q})$, which in turn further reduces q through (8). As before, on the other hand, p must stay at 0 in the new stationary equilibrium $(0, \tilde{q} - \Delta q)$.

In the stable B&C equilibrium $(1, \bar{q})$ the increase in λ reduces q as before, but the effect on the stigma cost may be uncertain due to the conflicting direct and indirect effects of λ. As stated before, the direct *negative* effect of λ dominates the indirect effect through varying q, thereby leading to an increased stigma cost and thus a decreasing q. We can explain the effects of changes in the wage rate in a manner similar to that in $(0, \tilde{q})$.

Besley and Coate (1992) 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, in general, **ambiguous** because a rise in benefits not only raises the fraction of undeserving claimants, $1 - \pi$, but also attracts more undeserving claimants with a lower disutility (i.e., $\bar{\theta}_u$ falls). Due to these two conflicting effects on $\bar{\theta}_w = \pi \bar{\theta}_d + (1 - \pi)\bar{\theta}_u = \theta_u$, the net effect on the stigma cost is uncertain. Although the effect on the stigma cost may be uncertain around $(1, \bar{q})$ (recall $s_q(1, \bar{q}) \gtreqless 0$ due to (13)) even in the present model, we see the *dominant* direct effect of changes in parameters such as b and λ on q. Consequently, we determine the effect on q unambiguously.

There are noteworthy policy implications that differ from those in Besley and Coate (1992). Besley and Coate (1992) are solely concerned with how to raise the stigma costs, thereby leading to a reduction in *welfare fraud*, while we are chiefly concerned with how to improve lower or non-take up rates. For this reason, we propose that policymakers should adopt the populations of deserving and undeserving claimants as policy targets rather than the level of stigma costs, unlike Besley and Coate (1992). Since the population of deserving claimants p remains unaffected in the long run, despite the changes in the structural parameters of the model or policy reforms, 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 would not. However, our replicator dynamic model would provide further policy implications if the stationary point $(1, \bar{q})$ is more socially desirable than $(0, \tilde{q})$ because there is no incomplete take-up in $(1, \bar{q})$.[6] From this prospective, there is room for government intervention that will make the long-run equilibrium $(1, \bar{q})$ more likely to occur. 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 realize Fig. 3 rather than Fig. 5. To this end, we need to reduce the intercept of the locus $\dot{p} = 0$ with the q-axis, denoted by $q^*|_{\dot{p}=0}$, while raising that of the locus $\dot{p} = 0$, denoted by $\tilde{q}|_{\dot{q}=0}$, where

$$q^*|_{\dot{p}=0} = 1 - [2u(b)/\lambda] \text{ and } \tilde{q}|_{\dot{q}=0} = 1 - \frac{2\,[v(\omega) - v(b)]}{2 - \lambda}. \tag{15}$$

Inspecting (15) reveals that higher levels of b shift the locus $q^*|_{\dot{p}=0}$ downwards, while shifting the locus $\tilde{q}|_{\dot{q}=0}$ upwards. Hence, we conclude that raising the level of welfare benefits is a more socially desirable policy than enhancing the degree of public exposure, either when policy makers are more concerned with the well-being of the deserving poor or when the social welfare evaluated at $(1, \bar{q})$ is higher than that in $(0, \tilde{q})$.

6 Tax-payer Resentment View Stigma

In this section, we introduce an alternative stigma costs function based on the tax-payer resentment view of welfare stigma suggested by Besley and Coate (1992) into the present replicator dynamic 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. Thus, we can express the government's budget constraint as:

$$(1 - \beta)T = b\beta\,[\gamma p + (1 - \gamma)q]\,, \tag{16}$$

[6]However, since the population of undeserving claimants in $(1, \bar{q})$ is larger than that in $(0, \tilde{q})$, it is a trade-off relation between the populations of undeserving and deserving claimants. Nevertheless, it seems that the problem of *non-take-up* in $(0, \tilde{q})$ is more serious.

where the left-hand side represents the total tax revenue and the right-hand side represents the cost for providing total welfare benefits. Thus, the level of the lump-sum tax must be equal to

$$T = \frac{b\beta\left[\gamma p + (1-\gamma)q\right]}{1-\beta}, \tag{17}$$

with the following properties:

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

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 Besley and Coate (1992), we assume not only that the rich individuals have different concerns (or degrees of compassion) for the deserving poor, measured by μ, so that the rich have different preferred benefit levels, but also that the preferences of the rich are $\log\left[y - T\right]$, where $y - T$ represents their private consumption and $P(b)$ is a measure of poverty or distress with $P'(b) < 0$.[7] 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:

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

where the second equality follows from the fact 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* (16).

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

$$\left[y - \frac{\beta(\gamma p + (1-\gamma)q)}{1-\beta} b\right]^{-1} \frac{\beta\left[\gamma p + (1-\gamma)q\right]}{1-\beta} = -\mu\beta\gamma P'(b). \tag{18}$$

The left-hand side of (18) represents the marginal utility cost of an increase in one unit of the tax payment, while the right-hand side represents the

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

marginal benefit of improving the well-being of the poor. To obtain a closed-form solution, we further assume $P(b) = -\eta b + c$ with a constant $\eta > 0$, whereby (18) simplifies to

$$\frac{\gamma p + (1-\gamma)q}{(1-\beta)(y-T)} = \mu\gamma\eta. \tag{19}$$

We solve (19) for b to obtain

$$b^*(\mu; p, q) = \frac{(1-\beta)y}{\beta\left[\gamma p + (1-\gamma)q\right]} - \frac{1}{\beta\gamma\eta\mu}. \tag{20}$$

Recalling that the respective rich individuals (i.e., tax-payers) have different values for μ, there certainly exists an individual having the *threshold weight* $\bar{\mu}$ 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.

$$b^*(\bar{\mu}; p, q) = b. \tag{21}$$

Consequently, we can decompose the remaining tax-payers 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 (21), coupled with (20), for $\bar{\mu}$ yields

$$\bar{\mu} = \frac{\gamma p + (1-\gamma)q}{\gamma\eta(1-\beta)(y-T)}. \tag{22}$$

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 appropriate or reasonable $b^*(\mu; p, q)$:

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

where the scale parameter of the stigma cost λ^r takes a constant and common value among the rich. As in Besley and Coate (1992), we define the stigma costs function as an increasing function of *aggregate* tax-payers' 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 (16), 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 (22) into the above expression and rearranging results in

$$s^r(p,q) = \frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{y-T}, \tag{23}$$

with the property

$$\frac{\partial s^r(p,q)}{\partial T} = \frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2} > 0. \tag{24}$$

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

6.1 Stability Properties of the Model of the Tax-payer Resentment View Stigma

The replicator dynamics is also described by (7) and (8), except that we replace the stigma cost function $s(p,q)$ by $s^r(p,q)$ in (23). 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. The following two lemmas serve in drawing the above two loci.

Lemma 1 *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 $(p,q)\in[0,1]^2$.*

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

$$\frac{dq}{dp} = -\frac{s^r_p(p,q)}{s^r_q(p,q)}.$$

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

$$\frac{dq}{dp} = -\frac{s^r_p(p,q)}{1+s^r_q(p,q)},$$

where it is straightforward to verify that

$$s_p^r(p,q) = \frac{\lambda^r}{2\beta\gamma\eta^2}\frac{1}{y-T} + \frac{\lambda^r}{2\gamma\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2}\frac{b}{1-\beta} > 0, \text{ and} \tag{25}$$

$$s_q^r(p,q) = \frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{1-\gamma}{y-T} + \frac{\lambda^r}{2\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2}\frac{b(1-\gamma)}{1-\beta} > 0. \tag{26}$$

Taken together, we have $\frac{s_p^r(p,q)}{s_q^r(p,q)} > \frac{s_p^r(p,q)}{1+s_q^r(p,q)}$. ■

Lemma 2 *The intercept of the locus $v(b) - v(\omega) + 1 - q - s^r(p,q) = 0$ on the vertical axis at 1, denoted by $\bar{q}^r$, is less than 1.*

Proof. We can rewrite the locus $v(b) - v(\omega) + 1 - q - s^r(p,q) = 0$, evaluated at the stationary point $(1, \bar{q}^r)$, as follows:

$$v(\omega) - v(b) = 1 - \bar{q}^r - s^r(1, \bar{q}^r). \tag{27}$$

Since the left-hand side of (27) is positive, the right-hand side is also positive. Moreover, since $s^r(1, \bar{q}^r) > 0$, $\bar{q}^r$ must be less than 1. ■

We summarize the stability properties of the respective stationary points as follows:

Proposition 3 *Under the replicator dynamics* (7) *and* (8)*, coupled with Assumptions 1, 2, and 3 being replaced by the stigma cost function* (23) *for all* $(p,q) \in [0,1]^2$*:*

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

(ii) The stationary point $(1, \bar{q}^r)$ is locally asymptotically stable 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 locally asymptotically stable 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 $(\bar{p}_2^r, 1)$ is a saddle.

(v) The stationary point $(\tilde{p}_2^r, 0)$ is locally asymptotically stable if $v(b) - v(\omega) + 1 - s^r(\tilde{p}_2^r, 0) < 0$. Conversely, if $v(b) - v(\omega) + 1 - s^r(\tilde{p}_2^r, 0) > 0$, then it is a saddle.

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

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

(ix) *The stationary point* $(0,0)$ *is either a saddle or a source.*

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

Proof. See Appendix B.[8] ∎

Based on Lemma 1 and Proposition 3, Figs. 6-9 provide several phase portraits for the dynamic behavior of p and q. We first consider Fig. 6. The stationary point $(1,\bar{q}^r)$ is located below the locus $\dot{p}=0$ (i.e., $\dot{p}>0$), implying that $u(b) - s^r(1,\bar{q}^r) > 0$, which therefore satisfies Claim (ii) of Proposition 3. Hence, it is locally asymptotically stable. However, the location of the stationary point $(0,\tilde{q}^r)$ below the locus $\dot{p}=0$ in Fig. 6 implies that $u(b) - s^r(0,\tilde{q}^r) > 0$. Thus, $(0,\tilde{q}^r)$ is unstable due to Claim (iii). The stationary point $(\bar{p}_2^r,1)$ is also unstable due to Claim (iv). Although the stationary point $(1,\bar{q}_2^r)$ is non-hyperbolic (because its determinant is equal to zero), case (x) in Appendix B shows that there is a trajectory starting from the initial condition located arbitrarily close to $(1,\bar{q}_2^r)$ that diverges from $(1,\bar{q}_2^r)$, implying that it is unstable. Other stationary points such as $(1,1)$, $(1,0)$, $(0,1)$, and $(0,0)$ are all unstable because they do not satisfy Claims (vi)-(ix), respectively. Taken together, we conclude that the only stationary point $(1,\bar{q}^r)$ is locally (eventually, globally) asymptotically stable in Fig. 6.

Fig. 7, on the other hand, has the *interior* stationary point $(\hat{p}^r,\hat{q}^r)$, which is locally asymptotically stable according to Claim (i), but all non-interior stationary points are unstable. In Fig. 8, the only stationary point $(0,\tilde{q}^r)$ is locally asymptotically stable because it satisfies Claim (iii) (noting that since it is located above the locus $\dot{p}=0$, $u(b) - s^r(0,\tilde{q}^r) < 0$ due to (7)), while the only stationary point $(\tilde{p}_2^r,0)$ in Fig. 9 is locally asymptotically stable because it satisfies Claim (v). Similarly, we can verify that all other stationary points in Figs. 8 and 9 are unstable. Notably, the only stationary point $(1,0)$ in Fig. 10 is locally asymptotically stable because it is located

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

below the locus $\dot{p} = 0$ (i.e., $u(b) - s^r(1,0) > 0$) and the locus $\dot{q} = 0$ (i.e., $v(b) - v(\omega) - s^r(1,0) < 0$); consequently, it satisfies Claim (*vii*).

Several remarks are in order. First, when there exists an *interior* stationary point such as in Fig. 7, it is always locally (i.e., eventually, globally) asymptotically stable. This feature stands in sharp contrast with the statistical discrimination view stigma model in which the interior stationary point $(\hat{p}^r, \hat{q}^r)$ is always *unstable*. Intuitively, in the tax-payer resentment model, if the initial populations of the deserving and undeserving claimants are too large (i.e., close to the stationary equilibrium $(1,1)$), then the rich have the burden of larger tax payments. Consequently, their resentment will be greatly intensified, thereby augmenting the stigma cost incurred by benefits claimants. Therefore, the higher stigma cost in turn discourages the deserving and undeserving claimants to take up welfare benefits. This leads to a reduction in their populations and thus prevents their populations from approaching $(1,1)$. On the contrary, when their populations are very much lower (i.e., close to $(0,0)$), then so is the stigma cost, thereby inducing more of the deserving and undeserving poor to take up welfare benefits, and thus preventing their populations from approaching $(0,0)$. These features indicate that there is always a force that leads the deserving and undeserving claimants to the interior stationary point $(\hat{p}^r, \hat{q}^r)$, as long as it exists. Second, if an interior stationary point does not exist, one of the stationary equilibria $(0,1)$ and $(1,0)$ may be stable depending on the relative strength of the preferences of the deserving and undeserving claimants towards welfare benefits (i.e., the relative magnitude of $u(b)$ and $v(b)$). More precisely, if $u(b) > s^r(1,0) > v(b) - v(\omega) + 1$, then the deserving claimants have relatively stronger preferences towards welfare benefits compared to the undeserving claimants, and thus the population of deserving claimants increases, although that of the undeserving claimants tends to fall due to the relatively large stigma cost. Consequently, $(1,0)$ will be stable, as illustrated in Fig. 10. Conversely, if $u(b) < s^r(0,1) < v(b) - v(\omega) + 1$, then the undeserving claimants have relatively stronger preferences towards welfare benefits, and thus the population of undeserving claimants rises, whereas that of deserving claimants declines because their utility from receiving benefits is less than the stigma cost. Thus, $(0,1)$ will be stable, as stated in Claim (*viii*).

6.2 Comparative Statics for Taxpayer Resentment View Stigma Model

This subsection conducts a comparative static analysis of the taxpayer resentment view stigma model. In particular, we analyze four stable stationary

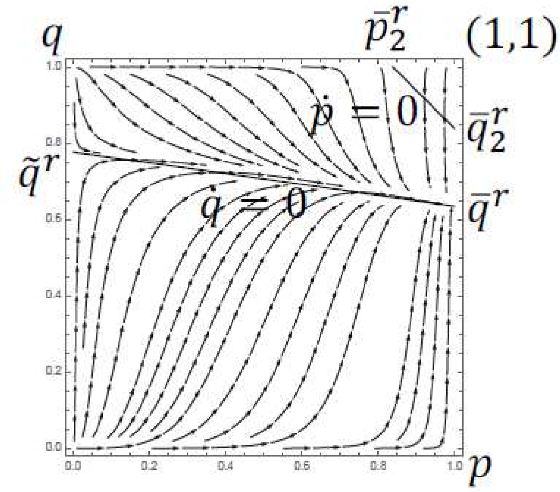

Figure 6: $\lambda^r = 0.3$, $u(b) = 0.3$, $v(b) = 0.4$, $v(\omega) = 0.5$, $\gamma = 0.5$, $\beta = 0.1$, $y = 1.2$, $b = 0.5$, and $\eta = 4$.

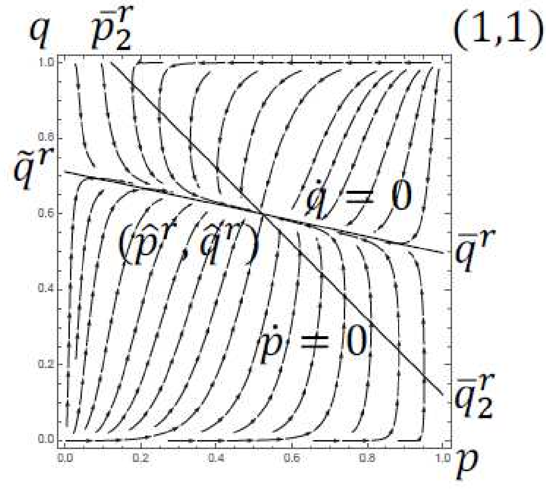

Figure 7: $\lambda^r = 0.5$, $u(b) = 0.3$, $v(b) = 0.4$, $v(\omega) = 0.5$, $\gamma = 0.5$, $\beta = 0.1$, $y = 1.2$, $b = 0.5$, and $\eta = 4$.

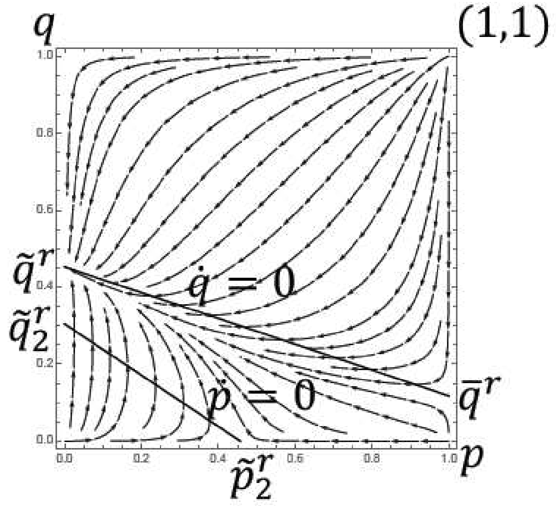

Figure 8: $\lambda^r = 1$, $u(b) = 0.3$, $v(b) = 0.4$, $v(\omega) = 0.5$, $\gamma = 0.4$, $\beta = 0.1$, $y = 1.2$, $b = 0.5$, and $\eta = 4$.

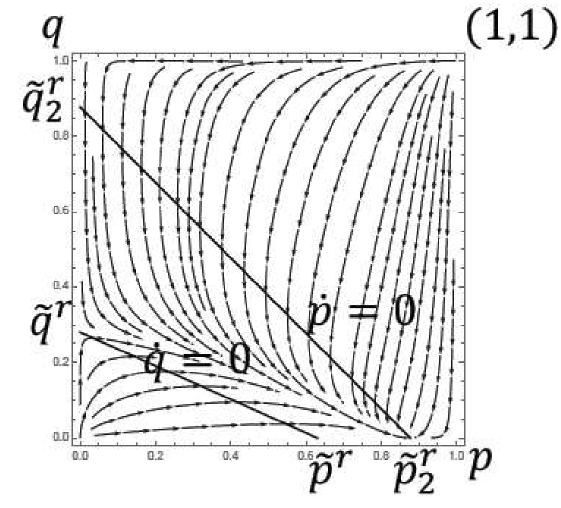

Figure 9: $\lambda^r = 1.5$, $u(b) = 0.7$, $v(b) = 0.4$, $v(\omega) = 0.9$, $\gamma = 0.5$, $\beta = 0.1$, $y = 1.2$, $b = 0.5$, and $\eta = 4$.

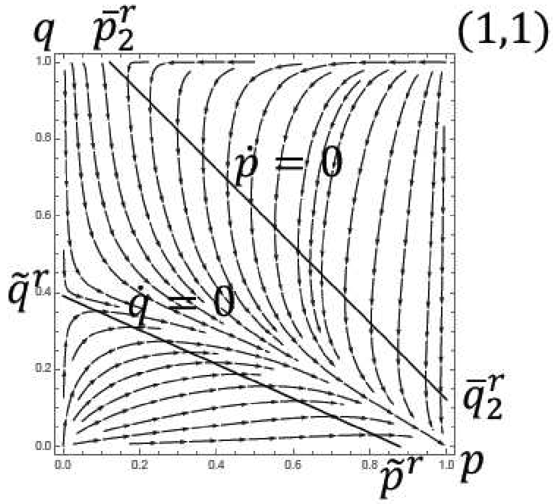

Figure 10: $\lambda^r = 1.5$, $u(b) = 0.9$, $v(b) = 0.2$, $v(\omega) = 0.5$, $\gamma = 0.5$, $\beta = 0.1$, $y = 1.2$, $b = 0.5$, and $\eta = 4$.

points: $(1, \bar{q}^r)$ in Fig. 6, $(\hat{p}^r, \hat{q}^r)$ in Fig. 7, $(0, \tilde{q}^r)$ in Fig. 8, and $(\tilde{p}_2^r, 0)$ in Fig. 9 with respect to changes in the parameters b, λ^r, y, ω, and η, respectively.[9]

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

$$\begin{aligned} u(b) - s^r(\hat{p}^r, \hat{q}^r) &= 0, \text{ and} && (28) \\ v(b) - v(\omega) + 1 - \hat{q}^r - s^r(\hat{p}^r, \hat{q}^r) &= 0, && (29) \end{aligned}$$

the above two conditions are reduced to

$$v(b) - v(\omega) + 1 - \hat{q}^r - u(b) = 0. \tag{30}$$

Differentiating (30) with respect to b yields

$$\frac{d\hat{q}^r}{db} = v'(b) - u'(b) \gtreqless 0, \tag{31}$$

while differentiating (28) with respect to b, coupled with (31), yields

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

[9] The stationary points $(1, 0)$ in Fig. 10 and $(0, 1)$ in Claim $(viii)$ of Proposition 3 are also locally asymptotically stable. Since these stationary points are unaffected by small changes in any parameter value, neverthless, we do not conduct a comparative statics analysis around them.

where $\partial \hat{s}^r(\hat{p}^r, \hat{q}^r)/\partial b = [\partial s^r(\hat{p}^r, \hat{q}^r)/\partial T]\,[\partial T(\hat{p}^r, \hat{q}^r)/\partial b] > 0$ due to $\partial s^r(\hat{p}^r, \hat{q}^r)/\partial T > 0$ from (24) and $\partial T(\hat{p}^r, \hat{q}^r)/\partial b = \beta\,[\gamma \hat{p}^r + (1-\gamma)\hat{q}^r]\,/\,(1-\beta) > 0$. In the stationary equilibrium $(1, \bar{q}^r)$, we differentiate (29) with respect to b to obtain

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

where $s^r_q(1, \bar{q}^r) > 0$ from (23) and

$$\frac{\partial s^r(1, \bar{q}^r)}{\partial b} = \frac{\partial s^r(1, \bar{q}^r)}{\partial T}\frac{\partial T(1, \bar{q}^r)}{\partial b} = \underbrace{\frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma + (1-\gamma)\bar{q}^r}{(y-T)^2}}_{(+)}\underbrace{\frac{\beta\,[\gamma + (1-\gamma)\bar{q}^r]}{1-\beta}}_{(+)} > 0.$$

In the stationary point $(0, \tilde{q}^r)$, differentiating (29) with respect to b and evaluating the result at $(0, \tilde{q}^r)$ yields

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

where $s^r_q(0, \tilde{q}^r) > 0$ from (23) and

$$\frac{\partial s^r(0, \tilde{q}^r)}{\partial b} = \frac{\partial s^r(0, \tilde{q}^r)}{\partial T}\frac{\partial T(1, \bar{q}^r)}{\partial b} = \underbrace{\frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{(1-\gamma)\tilde{q}^r}{(y-T)^2}}_{(+)}\underbrace{\frac{\beta\,[(1-\gamma)\tilde{q}^r]}{1-\beta}}_{(+)} > 0.$$

Finally, differentiating (28) with respect to b and evaluating the result at $(\tilde{p}^r, 0)$ yields

$$\frac{d\tilde{p}^r_2}{db} = \frac{u'(b) - [\partial s^r(\tilde{p}^r_2, 0)/\partial b]}{s^r_p(\tilde{p}^r_2, 0)} \gtreqless 0,$$

because $u'(b) - [\partial s^r(\tilde{p}^r_2, 0)/\partial b] \gtreqless 0$ and

$$\frac{\partial s^r(\tilde{p}^r_2, 0)}{\partial b} = \frac{\partial s^r(\tilde{p}^r_2, 0)}{\partial T}\frac{\partial T(\tilde{p}^r_2, 0)}{\partial b} = \underbrace{\frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma\tilde{p}^r_2}{(y-T)^2}}_{(+)}\underbrace{\frac{\beta\gamma\tilde{p}^r_2}{1-\beta}}_{(+)} > 0.$$

Similarly, we can carry out a comparative statics analysis with respect to the other parameters λ^r, γ, y, ω, and η. We summarize all of the results as follows (we provide the detailed mathematical derivations in Appendix C):

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 that on the population 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 also* ***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 also* ***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* $(\hat{p}^r, \hat{q}^r)$ *and* $(\tilde{p}_2^r, 0)$.

(v) *The effects of an increase in the efficacy of social welfare programs on the populations of deserving and undeserving claimants are* ***positive*** *in* $(\hat{p}^r, \hat{q}^r)$, $(1, \bar{q}^r)$*, 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 also* ***positive*** *in* $(\hat{p}^r, \hat{q}^r)$ *and* $(\tilde{p}_2^r, 0)$.

The economic interpretation is as follows. Let us 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, $b + \Delta b$, causes more of both the deserving and undeserving poor to take up benefit, which is implied by (7) and (8). Thus, the total tax burden unambiguously increases, which in turn strengthens the resentment of the tax-payers (i.e., the rich). Accordingly, the stigma cost is intensified, which in turn causes the populations of deserving and undeserving claimants to fall. Due to these conflicting effects on q, therefore, the net effect is ambiguous. In contrast, an increase in λ^r directly raises the stigma cost, thereby reducing the populations of deserving and/or undeserving claimants over time. The resulting decreased 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 dominates the *indirect negative* effect of the induced decrease in p and q. Consequently, the stigma cost unambiguously rises, thereby leading to a fall in p and/or q in the long run. However, since the stigma cost must be equal to $u(b)$ in the interior stationary point, q is determined solely according to (30), independently of λ^r. Hence, 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. Thus, the populations of deserving and/or undeserving claimants rise over time. However, since the stigma cost must be equal to $u(b)$ in the interior stationary equilibrium, the stationary value of q is determined solely by (30) as stated above, which is independent of y. In contrast, an increase in the wage rate, ω, depresses the population growth rate of the undeserving claimants because its increase induces them to work more. In the stationary equilibria $(\tilde{p}_2^r, 0)$ and $(\hat{p}^r, \hat{q}^r)$, however, the long run value of p is unaffected by changes in ω because in the stationary equilibrium the stigma cost $s(\tilde{p}_2^r, 0)$ must be equal to $u(b)$ and (30) does not depend on ω. Finally, enhancing the efficacy of social welfare programs (i.e., increasing μ) in general raises the population growth rates of the deserving and/or undeserving claimants. This occurs because the enhanced efficacy makes the rich more generous, which in turn reduces the stigma cost. Thus, the populations of undeserving and/or deserving claimants tend to rise for the reason stated above, except in the interior stationary equilibrium.

7 Concluding Remarks

We investigated the case in which welfare fraud and incomplete take-up *simultaneously* emerge, two phenomena of apparently opposite nature, in a simple setting of replicator dynamics, which allows the *heterogeneous* populations of deserving and undeserving claimants to endogenously evolve over time, unlike the *static* model of Besley and Coate (1992). This study found multiple stationary equilibria in the replicator dynamics models of the statistical-discrimination view stigma as well as the tax-payer resentment view stigma suggested by Besley and Coate (1992). We applied a stability analysis to reduce the set (or multiplicity) of possible long-run equilibria substantially, allowing us to sharpen the predictive power of the comparative statics analysis by ruling out unstable stationary equilibria. Notably, there may coexist at most two stable stationary equilibria in the statistical discrimination view stigma model. In this case, depending on the initial proportions of deserving and undeserving claimants, different long-run outcomes will emerge. More importantly, different stationary equilibria deliver different policy implications. In particular, there is no interior **stable** stationary equilibrium in the statistical discrimination view stigma model, which implies that welfare fraud and incomplete take-up never *simultaneously* emerge in the long run; instead, one of the boundary stationary equilibria with 100% incomplete take-up and with 100% take-up exclusively does.

In contrast, in the tax-payer resentment view stigma model, the interior stationary equilibrium is globally asymptotically stable, provided it exists. Hence, irrespective of the initial proportions of deserving and undeserving claimants, the coexistence of those claimants will be present in the long-run. In other words, such coexistence is more likely to emerge when stigma costs arise from the tax-payers' (or the rich) resentment. 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 tax-payer resentment view stigma model relies on the **total population of claimants**. In the former model, if the welfare fraud population is large, then the deserving poor who want to receive welfare benefits are discouraged by higher stigma costs. Ultimately, all of the deserving poor end up giving up benefits in the long run. On the contrary, if the initial welfare fraud population is small, then all deserving poor end up taking up benefits in the long run. These extremely long-run outcomes stem from the cumulative spiral process, which operates such that the deserving claimants with incomplete take-up tend to enforce themselves, thus leading to either 100% incomplete or 100% complete take-up in the long run. In other words, this somewhat destabilizing feature stems from the divergent property of the average disutility of welfare claimants. In contrast, the tax-payer resentment model displays a relatively more stabilizing tendency towards an interior stationary equilibrium that allows for the coexistence of welfare fraud and incomplete take-up. This tendency stems from the fact that aggregate externalities arising from the total population of welfare claimants, rather than the composition of heterogeneous populations, matters in determining the size of stigma costs. The latter model produces long-run outcomes that would better fit actual observations in Japan, Germany, and the like.

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 that between the poor and the rich are 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. To relax the assumption of the fixed share of populations, we may introduce stochastic perturbations or random shocks that reflect a random switch from the deserving poor to the undeserving poor, or vice versa, or a random switch from the poor to the rich, or vice versa. We could utilize the perturbed replicator dynamic model of Samuelson and Zhang (1992), in which agents are subject to stochastic perturbations, for such an analysis. Even in the tax-payer resentment model, the government simply imposes lump-sum taxes and transfers those tax revenues to welfare claimants. However, this scheme is obviously unrealistic because such transfers are mainly financed by distortionary taxes in most countries. Nevertheless, by introducing distortionary taxes, the government has to face incentive compatibility constraints to cope with the accompanying asymmetric information problem, as in Blumkin et al. (2015). An optimal welfare program solution should be part of *non-linear* optimal income taxation, as in the Mirrlees (1971) framework, which is augmented to allow for asymmetric information on tax-payers' earning ability and for the extended margin choices of the undeserving poor. Alternatively, 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 parameter choices into our replicator dynamics framework.

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

Appendices

Appendix A: Stability Analysis

(i) The determinant of the Jacobian for the linearized system (11), which is in evaluated at the stationary point $(1,\bar{q})$, is given by

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

where it follows from (13) that

$$s(1,\bar{q}) = \frac{\lambda}{2}\frac{(1-\gamma)\bar{q}\,(1-\bar{q})}{\gamma+(1-\gamma)\bar{q}} > 0 \text{ and} \tag{A.2}$$

$$\frac{\partial s(1,\bar{q})}{\partial q} = \frac{\lambda}{2}\frac{(1-\gamma)\left[\gamma(1-2\bar{q})-(1-\gamma)\bar{q}^2\right]}{\left[\gamma+(1-\gamma)\bar{q}\right]^2} \gtreqless 0. \tag{A.3}$$

By substitution of (A3), moreover, it can be confirmed that

$$1+\frac{\partial s(1,\bar{q})}{\partial q} = \frac{\gamma^2+\frac{\lambda}{2}(1-\gamma)\gamma+(1-\gamma)\bar{q}(1-\frac{\lambda}{2})\left[2\gamma+(1-\gamma)\bar{q}\right]}{\left[\gamma+(1-\gamma)\bar{q}\right]^2} > 0. \tag{A.4}$$

Taken together, it turns out that the sign of the determinant (A.1) is determined according to that of the term $u(b)-s(1,\bar{q})$. Hence, if $u(b)-s(1,\bar{q})>0$, then (A.1) is positive, and the trace is negative. These facts imply that $(1,\tilde{q})$ is asymptotically locally stable. Conversely, if $u(b)-s(1,\bar{q})<0$ holds, then (A.1) is negative and thus $(1,\bar{q})$ is a saddle.

(ii) The determinant evaluated at the stationary point $(0,\tilde{q})$ is given by

$$\begin{aligned} |J|_{(0,\tilde{q})} &= \begin{vmatrix} u(b)-s(0,\tilde{q}) & 0 \\ -\tilde{q}(1-\tilde{q})s_p(0,\tilde{q}) & -\tilde{q}(1-\tilde{q})\left[1+s_q(0,\tilde{q})\right] \end{vmatrix}, \\ &= -\left[u(b)-s(0,\tilde{q})\right]\tilde{q}(1-\tilde{q})\left[1+s_q(0,\tilde{q})\right], \end{aligned} \tag{A.5}$$

where it follows from (6) that $s(0,\tilde{q})=(\lambda/2)(1-\tilde{q})>0$, while it follows from (13) that

$$1+\frac{\partial s(0,\tilde{q})}{\partial q}=1+\frac{\lambda}{2}(1-\tilde{q})>0. \tag{A.6}$$

Taken together, if $u(b)-s(0,\tilde{q})>0$, then the determinant (A.5) is negative, which implies that $(0,\tilde{q})$ 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 asymptotically locally stable.

(iii) The determinant evaluated at the interior stationary point $(\hat{p},\hat{q})$ is given by

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

which implies that it is a saddle.

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

$$|J|_{(1,1)} = \begin{vmatrix} -u(b) & 0 \\ 0 & v(\omega)-v(b) \end{vmatrix} = -u(b)\left[v(\omega)-v(b)\right]<0, \tag{A.7}$$

$$|J|_{(0,1)} = \begin{vmatrix} u(b) & 0 \\ 0 & -\left[v(b)-v(\omega)\right] \end{vmatrix} = u(b)\left[v(\omega)-v(b)\right]>0, \tag{A.8}$$

$$|J|_{(1,0)} = \begin{vmatrix} -u(b) & 0 \\ 0 & v(b)-v(\omega)+1 \end{vmatrix} = -u(b)\left[v(b)-v(\omega)+1\right]. \tag{A.9}$$

The determinant (A.7) is negative since $v(\omega)-v(b)>0$, implying that $(1,1)$ is a saddle. Since $v(\omega)-v(b)>0$, the determinant (A. 8) and the trace both are positive, which implies that $(0,1)$ is a saddle. The determinant evaluated at the stationary point $(1,0)$ is ambiguous. If $v(b)-v(\omega)+1<0$, the determinant (A. 8) is positive and the trace is negative, which together implies that is locally asymptotically stable. If $v(b)-v(\omega)+1>0$, (A. 9) is negative, which implies that $(1,0)$ is a saddle.

(v) The determinant evaluated at the stationary point $(0,\tilde{q}_2)$ is given by

$$|J|_{(0,\tilde{q}_2)}=\begin{vmatrix} 0 & 0 \\ -\tilde{q}_2(1-\tilde{q}_2)s_p(0,\tilde{q}_2) & \begin{array}{c}(1-2\tilde{q}_2)\left[v(b)-v(\omega)+1-\tilde{q}_2-s(0,\tilde{q}_2)\right] \\ +\tilde{q}_2(1-\tilde{q}_2)\left[1+s_p(0,\tilde{q}_2)\right]\end{array} \end{vmatrix}=0,$$

which implies that $(0,\tilde{q}_2)$ is *non-hyperbolic*; consequently, we cannot apply the linear approximation technique. Alternatively, consider any trajectory whose initial condition is located in a sufficiently small neighborhood of $(0,\tilde{q}_2)$ but above the curve $\dot{p}=0$. This trajectory entails $\dot{p}>0$ and $\dot{q}>0$ in the region denoted by Region I in Fig.3 (we consider only Fig. 3 because the almost parallel argument can be applied to the case of $(0,\tilde{q}_2)$ in Figs. 4 and 5).[10] As a result, this trajectory may cross the locus $\dot{q}=0$ or reach $(0,\bar{q})$. Suppose that this trajectory crosses the locus $\dot{q}=0$ at time t^*; hence, it holds that $dq(t^*)/dt=0$. Since it follows from $dp(t^*)/dt>0$ in Region I and (12) that

$$\frac{d^2q(t^*)}{dt^2}=-q(t^*)(1-q(t^*))\frac{\partial s(p(t^*),q(t^*))}{\partial p}\frac{dp(t^*)}{dt}>0,$$

$q(t)$ takes a minimum value at t^*. However, it is impossible because $q(t)$ always increases in this region. Hence, any trajectory starting from the initial condition located in the above-mentioned neighborhood of $(\hat{p}^r,\hat{q}^r)$ remains forever in this region. Moreover, since $p(t)$ monotonically decreases and $q(t)$ monotonically increases in this region, $p(t)$ and $q(t)$ converge the upper and lower limiting points, respectively; thus (p^r,q^r) converges $(0,\bar{q})$. As a result, we conclude that $(0,\bar{q})$ is **not** asymptotically locally stable.[11]

[10] In Fig. 5 there are two more stationary points. It is straightforward to show that the determinants evaluated at either stationary point are equal to zero, implying that both stationary points are non-hyperbolic. In a way similar to **(v)** we can show that neither stationary point is locally asymptotically stable.

[11] The differential equations of (7) and (8) are not defined at $(0,0)$ because of the discontinuity of the function $s\,(p,q)$ at that point. To get a rough idea regarding their dynamic behavior around a small neighborhood of the stationary point $(0,0)$, we may perturb $s\,(p,q)$ by a sufficiently small $\varepsilon>0$ such that

$$s(p,q;\varepsilon)=\frac{\lambda}{2}\frac{(1-\gamma)q(1-q)}{\gamma p+(1-\gamma)q+\varepsilon},\ \text{with } s(0,0;\varepsilon)=0,$$

Appendix B: Proof of Proposition 3

(i) The determinant of the tax resentment-view model evaluated at the interior stationary point $(\hat{p}^r, \hat{q}^r)$ is given by:

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

while the trace is given by $-\hat{p}^r(1-\hat{p}^r)s_p^r(\hat{p}^r,\hat{q}^r)-\hat{q}^r(1-\hat{q}^r)\left[1+s_q^r(\hat{p}^r,\hat{q}^r)\right] < 0$, which follows from $s_q^r(\hat{p}^r,\hat{q}^r) > 0$ (recall (26)). These facts together imply that $(\hat{p}^r,\hat{q}^r)$ is locally asymptotically stable.

(ii) The determinant evaluated at the stationary point $(1,\bar{q}^r)$, which is the intersection between the locus $\dot{q}=0$ and $p=1$, is given by:

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

Since $s_q^r(1,\bar{q}^r) > 0$ from (26), the sign of $|J^r|_{(1,\bar{q}^r)}$ is determined according to that of $u(b)-s^r(1,\bar{q}^r)$. If $u(b)-s^r(1,\bar{q}^r) > 0$, $|J|_{(1,\bar{q}^r)} > 0$ and the trace is given by $-[u(b)-s^r(1,\bar{q}^r)]-\bar{q}^r(1-\bar{q}^r)\left[1+s_q^r(1,\bar{q}^r)\right] < 0$. Taken together, it is locally asymptotically stable. If $u(b)-s^r(1,\bar{q}^r) < 0$, the determinant is negative and thus it is a saddle.

(iii) The determinant evaluated at the stationary point $(0,\tilde{q}^r)$, which is the intersection between the locus $\dot{q}=0$ and $p=0$, is given by:

$$s_p(0,0;\varepsilon) = -\frac{\lambda}{2}\frac{\gamma(1-\gamma)q(1-q)}{[\gamma p+(1-\gamma)q+\varepsilon]^2}\bigg|_{p=q=0} = 0 \text{ and } \frac{\partial s(p,q;\varepsilon)}{\partial q}\bigg|_{p=q=0} = \frac{\lambda(1-\gamma)}{2}\frac{1}{\varepsilon} > 0.$$

In this case, the determinant evaluated at $(0,0)$ is given by

$$\begin{aligned} |J|_{(0,0)} &= \begin{vmatrix} u(b)-s(0,0;\varepsilon)-0\cdot s_p(0,0;\varepsilon) & 0\cdot s_q(0,0;\varepsilon) \\ 0\cdot s_p(0,0;\varepsilon) & \begin{array}{c} v(b)-v(\omega)+1-s(0,0;\varepsilon) \\ +0\cdot[1+s_q(0,0;\varepsilon)] \end{array} \end{vmatrix}, \\ &= u(b)\left[v(b)-v(\omega)+1\right] \gtreqless 0, \end{aligned}$$

which implies that if $v(b)-v(\omega)+1 < 0$, $(0,0)$ is a saddle, while if $v(b)-v(\omega)+1 > 0$, it is a source.

$$
\begin{aligned}
|J^r|_{(0,\tilde{q}^r)} &= \begin{vmatrix} u(b) - s^r(0,\tilde{q}^r) & 0 \\ -\tilde{q}^r(1-\tilde{q}^r)s^r_p(0,\tilde{q}^r) & -\tilde{q}^r(1-\tilde{q}^r)\left[1+s^r_q(0,\tilde{q}^r)\right] \end{vmatrix}, \\
&= -[u(b) - s^r(0,\tilde{q}^r)]\tilde{q}^r(1-\tilde{q}^r)\left[1+s^r_q(0,\tilde{q}^r)\right],
\end{aligned}
$$

while the trace is given by $u(b) - s^r(0,\tilde{q}^r) - \tilde{q}^r(1-\tilde{q}^r)\left[1+s^r_q(0,\tilde{q}^r)\right]$, where $s^r(0,\tilde{q}^r) > 0$ and thus $1 + s^r_q(0,\tilde{q}^r) > 0$. Hence, if $u(b) - s^r(0,\tilde{q}^r) < 0$, then $|J|_{(0,\tilde{q}^r)} > 0$ and the trace is negative, thus implying that $(0,\tilde{q}^r)$ is locally asymptotically stable. If $u(b) - s^r(0,\tilde{q}^r) > 0$, then $|J|_{(0,\tilde{q}^r)} < 0$, thus implying that it is a saddle.

(iv) The determinant evaluated at the stationary point $(\bar{p}^r_2, 1)$, which is the intersection between the locus $\dot{p} = 0$ and $q = 1$, is given by:

$$
\begin{aligned}
|J^r|_{(\bar{p}^r_2,1)} &= \begin{vmatrix} -\bar{p}^r_2(1-\bar{p}^r_2)s^r_p(\bar{p}^r_2,1) & -\bar{p}^r_2(1-\bar{p}^r_2)s^r_q(\bar{p}^r_2,1) \\ 0 & -\left[v(b) - v(\omega) - s^r(\bar{p}^r_2,1)\right] \end{vmatrix}, \\
&= \bar{p}^r_2(1-\bar{p}^r_2)s^r_p(\bar{p}^r_2,1)\left[v(b) - v(\omega) - s^r(\bar{p}^r_2,1)\right] < 0,
\end{aligned}
$$

whose negative sign is due to $v(b) < v(\omega)$, which implies that $(\bar{p}^r_2, 1)$ is a saddle.

(v) The Jacobian evaluated at the stationary point $(\tilde{p}^r_2, 0)$, which is the intersection between the locus $\dot{p} = 0$ and $q = 0$, is given by:

$$
\begin{aligned}
|J^r|_{(\tilde{p}^r_2,0)} &= \begin{vmatrix} -\tilde{p}^r_2(1-\tilde{p}^r_2)s^r_p(\tilde{p}^r_2,0) & -\tilde{p}^r_2(1-\tilde{p}^r_2)s^r_q(\tilde{p}^r_2,0) \\ 0 & v(b) - v(\omega) + 1 - s^r(\tilde{p}^r_2,0) \end{vmatrix}, \\
&= -\tilde{p}^r_2(1-\tilde{p}^r_2)s^r_p(\tilde{p}^r_2,0)\left[v(b) - v(\omega) + 1 - s^r(\tilde{p}^r_2,0)\right] \gtreqless 0,
\end{aligned}
$$

while the trace is $-\tilde{p}^r_2(1-\tilde{p}^r_2)s^r_p(\tilde{p}^r_2,0) + v(b) - v(\omega) + 1 - s^r(\tilde{p}^r_2,0)$, where $s^r_p(\tilde{p}^r_2,0) > 0$ from (26). If $v(b) - v(\omega) + 1 - s^r(\tilde{p}^r_2,0) < 0$, $|J^r|_{(\tilde{p}^r_2,0)} > 0$ and the trace is negative, thus implying that $(\tilde{p}^r_2, 0)$ is locally asymptotically stable. Conversely, if the opposite sign holds, $|J^r|_{(\tilde{p}^r_2,0)} < 0$ and thus it is a saddle.

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

$$
\begin{aligned}
|J^r|_{(1,1)} &= \begin{vmatrix} -[u(b) - s^r(1,1)] & 0 \\ 0 & -\left[v(b) - v(\omega) - s^r(1,1)\right] \end{vmatrix}, \\
&= [u(b) - s^r(1,1)][v(b) - v(\omega) - s^r(1,1)],
\end{aligned}
$$

while the trace is given by $-[u(b) - s^r(1,1)] - [v(b) - v(\omega) - s^r(1,1)]$, where $s^r(1,1) > 0$. As a result, $(1,1)$ is locally asymptotically stable if $u(b) -$

$s^r(1,1) > 0$ and $v(b) - v(\omega) - s^r(1,1) > 0$. Otherwise, it is a source or a saddle.

(vii) The determinant evaluated at the stationary point $(1,0)$ is given by:

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

while the trace is given by $-[u(b) - s^r(1,0)] + v(b) - v(\omega) + 1 - s^r(1,0)$. If $u(b) - s^r(1,0) > 0$ and $v(b) - v(\omega) + 1 - s^r(1,0) < 0$, $|J^r|_{(1,0)}$ is positive and the trace is negative, thus implying that $(1,0)$ is locally asymptotically stable. Otherwise, it is either a source or a saddle.

(viii) The determinant evaluated at the stationary point $(0,1)$ is given by

$$
\begin{aligned}
|J^r|_{(0,1)} &= \begin{bmatrix} u(b) - s^r(0,1) & 0 \\ 0 & -[v(b) - v(\omega) + 1 - s^r(0,1] \end{bmatrix}, \\
&= -[u(b) - s^r(0,1)][v(b) - v(\omega) + 1 - s^r(0,1)] ,
\end{aligned}
$$

while the trace is given by $u(b) - s^r(0,1) - [v(b) - v(\omega) + 1 - s^r(0,1]$ with $s^r(0,1) > 0$. If $u(b) - s^r(0,1) < 0$ and $v(b) - v(\omega) + 1 - s^r(0,1) > 0$, $(0,1)$ is locally asymptotically stable. Otherwise, it is either a saddle or a source.

(ix) The determinant evaluated at $(0,0)$ is given by:

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

due to $s^r(0,0) = 0$, while the trace is given by $u(b) + v(b) - v(\omega) + 1$. If $v(b) - v(\omega) + 1 > 0$, the signs of $|J^r|_{(1,0)}$ and the trace are positive and thus $(0,0)$ is a source. Conversely, if the opposite sign holds, it is a saddle.

(x) Since it is immediate that the determinants evaluated at the stationary points $(1, \bar{q}_2^r)$, $(0, \tilde{q}_2^r)$ and $(\tilde{p}^r, 0)$ are all equal to zero, we cannot apply the linear approximation technique to either case. Following the same procedure as in case (**v**) of Appendix A, we can prove that there exists a sufficiently small neighborhood of $(1, \bar{q}_2^r)$ which lies above the locus $\dot{p} = 0$ in Figs. 7, 8, and 9 such that if any trajectory starts from the initial condition lying in this region, it never converges to $(0, \bar{q}_2^r)$. In a similar fashion, we can prove nonconvergence to $(0, \tilde{q}_2^r)$ and $(\tilde{p}^r, 0)$ as well.

Appendix C: Comparative Statics

Since (30) does not contain the parameter λ^r at the interior stationary point $(\hat{p}^r, \hat{q}^r)$, we have $d\hat{q}^r/d\lambda^r = 0$. Using this result, we differentiate (28) with respect to b to yield

$$\frac{d\hat{p}^r}{d\lambda^r} = -\frac{1}{s_p^r(\hat{p}^r,\hat{q}^r)}\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial \lambda^r} < 0,$$

since $\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial \lambda^r} = \frac{1}{2\beta\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{y-T} > 0$. In the other stationary points such as $(1,\bar{q}^r)$, $(0,\tilde{q}^r)$, and $(\tilde{p}_2^r, 0)$, we have

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

since it follows from (23) that $\partial s^r(\cdot)/\partial\lambda^r > 0$.

Next, since (30) does not depend on the income level of the rich, y, we have $d\hat{q}^r/dy = 0$, while differentiating (28) with respect to y yields

$$\frac{d\hat{p}^r}{dy} = -\frac{1}{s_p^r(\hat{p}^r,\hat{q}^r)}\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial y} > 0,$$

since $\frac{\partial s^r(\hat{p}^r,\hat{q}^r)}{\partial y} = -\frac{\lambda^r}{2\beta\gamma^2\eta^2}\frac{\gamma p+(1-\gamma)q}{(y-T)^2} < 0$. Similarly, in other stationary points the effects of changes in y are given by:

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

Differentiating (30) with respect to ω yields $d\hat{q}^r/d\omega = -v^{'}(\omega) < 0$, while differentiating (28) with respect to ω yields $d\hat{p}^r/d\omega = 0$, because $s^r(\hat{p}^r,\hat{q}^r)$ in (23) does not depend on the wage rate ω. Similarly, in other stationary equilibria we have:

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

where noting that $s^r(\hat{p}^r, \hat{q}^r)$ in (23) does not depend on ω.

Finally, differentiating (23) with respect to η results in

$$\frac{\partial s^r}{\partial \eta} = -\frac{\lambda^r}{\beta\gamma^2\eta^3}\frac{\gamma p + (1-\gamma)q}{y-T} < 0.$$

Since (30) does not contain η, in the interior stationary point $(\hat{p}^r, \hat{q}^r)$, we have $d\hat{q}^r/d\eta = 0$, while differentiating (28) with respect to η yields

$$\frac{d\hat{p}^r}{d\eta} = -\frac{1}{s_p^r(\hat{p}^r, \hat{q}^r)}\frac{\partial s^r(\hat{p}^r, \hat{q}^r)}{\partial \eta} > 0.$$

Similarly, in the other stationary equilibria we have

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

where it follows from (23) that $\partial s^r(\cdot)/\partial \eta < 0$.

References

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

Bhargava, S. and Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, 105(11), 3489-3529.

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.

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.

Larry Samuelson, L. and Zhang, J. (1992). Evolutionary Stability in Asymmetric Games. *Journal of Economic Theory*, 57(2), 363-391.

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. (2001). Rational poverty or poor rationality? The take-up of social assistance benefits. *Review of Income and Wealth*, 47(3), 379-398.

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

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