Tipping points and the dynamics of labour market segregation in Germany*

Sébastien Willis†

February 2022

Abstract

Tipping points in the composition of groups have been argued to explain observed patterns of segregation in residential markets, labour markets, or schools. I use administrative data from Germany from 1975–2010 to study whether such tipping dynamics are observed in the composition by nativity of firms and local industries. I reject the existence of tipping points in both cases. Furthermore, confidence intervals constructed using methods robust to small effects are quite narrow. My findings imply that, unlike other forms of segregation, workplace segregation is unlikely to be explained by the preferences of workers over the composition of their workplaces.

Keywords: Segregation, tipping points, immigration
JEL codes: J15, J61

*An earlier version of this paper was the first chapter of my dissertation at Universitat Pompeu Fabra. I would like to thank my advisors, Albrecht Glitz and Ruben Enikolopov, as well as Caterina Alacevich, Lídia Farré, Flavio Hafner, Linna Martén, Joan Monras, Dávid Nagy, Ana Rodríguez González, Matti Sarvimäki, and seminar participants at the UPF applied seminar, CREI international lunch, and Uppsala University for helpful feedback. Thanks also to Adam McCloskey for helpful discussions about implementing the inference methods in Andrews et al. (2021) and to Heinier Frank and the staff at the IAB for help accessing the data.

†Department of Economics, Uppsala University, sebastien.willis@nek.uu.se. This paper uses data from the Establishment History Panel 1975–2019, v2 (doi: 10.5164/IAB.BHP7519.de.en.v2), available upon application to the Institute for Employment Research. There are no relevant conflicts of interest or sources of funding to declare.

1 Introduction

Workplace segregation by nativity pervades the labour market in many developed countries.[1] In Germany, where the foreign-born made up 12.8 per cent of the population in 2008 (OECD, 2020), 40 per cent of immigrants would have needed to change firms to achieve a degree of segregation consistent with a random assignment of workers to firms (Glitz, 2014).[2] However, while there is ample cross-sectional evidence of segregation in the workplace, there is to date no empirical evidence on the dynamics of nativity-based workplace segregation. This is in contrast to other settings where segregation has been documented, and in particular residential neighbourhoods. There, the composition of neighbourhoods has been shown to follow a so-called tipping dynamic (Aldén et al., 2015; Card et al., 2008, 2011).

In this paper I study whether such tipping dynamics also exist in the composition of workplaces. An example of a tipping dynamic is given in Figure 1. For low values of the immigrant share, native workforce growth is unrelated to the base-year immigrant share. However, should the immigrant share cross some threshold, labelled s^* and referred to as a tipping point, native workforce growth falls sharply, potentially becoming negative, as the the firm's hiring shifts towards immigrants. The existence of such a discontinuity for some value of the base-year immigrant share is the constitutive characteristic of tipping dynamics and is the key empirical regularity I will test for in this paper.

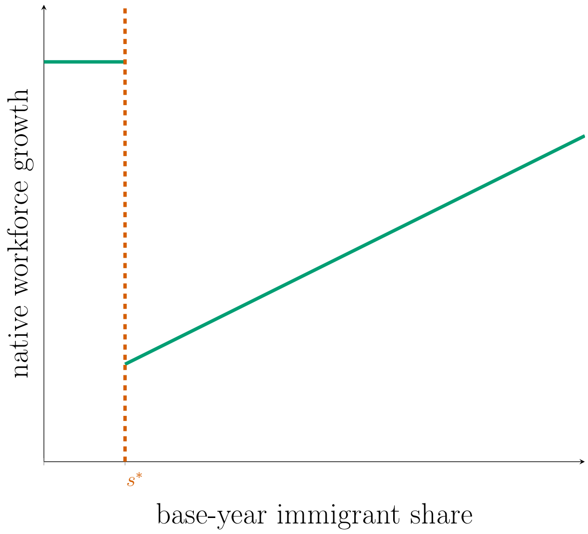
Prior work on residential segregation has related the existence of tipping points to the idea of preference spillovers—natives may prefer to live around other natives (Becker and Murphy, 2000; Card et al., 2008; Schelling, 1971, 1978). Formally, the existence of tipping points has been shown to be a necessary condition for there to be preference spillovers in the choice of residential location. While the existence of tipping points in the composition of workplaces is ultimately an empirical question, there are grounds to believe that similar preference spillovers may exist in the labour market. In 2017, only 37 per cent of Germans stated they would be "totally comfortable" having an immigrant as a work colleague, similar to the proportion (36 per cent) stating that they would be totally comfortable having an immigrant as a neighbour (European Commission, 2018).[3]

[1] Examples of countries where workplace segregation has been documented are the US (Hellerstein and Neumark, 2008; Andersson et al., 2014), Germany (Glitz, 2014), or Sweden (Åslund and Skans, 2010).

[2] Workplace segregation unexplained by observed characteristics suggests factors of production are misallocated, which could have large negative consequences for aggregate productivity and output (Hsieh et al., 2019). At the individual level, segregation across workplaces or industries could help explain the widely-studied persistence of employment and wage gaps between immigrants and natives (e.g. Lubotsky, 2007; Sarvimäki, 2011) and the fact that immigrants tend to work at lower-paying firms (Aydemir and Skuterud, 2008; Barth et al., 2012).

[3] The other options were "somewhat comfortable", "somewhat uncomfortable", "totally uncomfortable", or "don't know". Across the EU, the share "totally comfortable" was 43 per cent for neighbours and 44 per cent for colleagues.

Figure 1: Example of tipping dynamics



Furthermore, while the dynamics of segregation in labour markets have in general been less studied than residential markets, Pan (2015) has found evidence of tipping points in the gender composition of occupations, suggesting the mechanisms that explain segregation in the labour market could be similar to those underlying residential segregation.[4]

I test for tipping points in the composition of both firms and narrowly defined industries operating within a given local labour market in Germany, over the period 1975–2010, which I divide into five-year subperiods. Formally, the test for the presence of tipping points takes the form of a test for a downward intercept shift in native workforce growth net of immigrant workforce growth as a function of the base-year immigrant share. For both units of observation, I reject the existence of tipping points in the composition of workplaces. Furthermore, my estimates are relatively precise; the median lower bound of a 95 per cent confidence interval for the intercept shift rules out downward shifts larger than around 15 percentage points in the case of firms. This magnitude is equal to approximately 15–30 per cent of the magnitude of the discontinuities documented by Pan (2015) in the case of occupations.

Testing for the presence of tipping points presents econometric challenges, since the location of the tipping point is unknown to the researcher. Tests for tipping points in other settings have typically followed the method proposed by Card et al. (2008, 2011), which treats finding the location of the tipping point and testing for a discontinuity at the tipping point as separate problems. However, such a method may not be appropriate if the discontinuity is small relative to sampling variation, which will be the case if there is in fact no discontinuity at all. I therefore simultaneously identify the location of the

[4]In unpublished work, Zheng (2014) finds some evidence of tipping points in the racial composition of firms in the US, although she does not study the nativity of workers, nor are her methods robust to the null of no tipping point being true.

tipping point and the size of the discontinuity via nonlinear least squares and use the methods proposed by Andrews et al. (2019, 2021) for conducting inference on the size of a discontinuity when the location of the discontinuity is unknown. Since this paper is the first I am aware of to apply the inference methods developed by Andrews et al. (2021), I detail how these are applied in my setting in an online appendix.

The methods developed by Andrews et al. (2021), like those proposed by Card et al. (2008), assume that the location of the tipping point is common to the group of firms for which one is testing for a tipping point. However, if there are firm-specific amenities that matter differently to natives and immigrants, the location of the tipping point might be specific to each firm (Banzhaf and Walsh, 2013; Caetano and Maheshri, 2017). If firms differ sufficiently in the location of their firm-specific tipping point, a method that assumes a common tipping point will fail to find evidence of tipping dynamics, which could explain my null result.

I therefore consider numerous robustness checks that group firms in ways that proxy for unobserved amenities. In particular, I consider grouping firms by their fixed effect in an individual wage regression, consistent with recent evidence on the role of firm-level amenities in driving variation in the firm-specific component of wages (Sorkin, 2018), as well as by narrowly defined industry, and repeat the test for tipping dynamics. The estimated discontinuities in these cases are still centred on zero, even if they are a little less precisely estimated. When grouping firms by wage fixed effects, for example, the median lower bound of a 95 per cent confidence interval is around 25 percentage points; I continue to reject the existence of tipping points in the composition of firms when considering alternate groupings of firms.

Given that tipping points are a necessary condition for the presence of preference spillovers, the evidence presented here for native workforce growth being a smooth function of the base-year immigrant share suggests that workers do not have strong preferences over the composition of their set of coworkers. Preference spillovers are therefore unlikely to explain observed patterns of workplace segregation. The causes of the segregation of immigrants from natives across workplaces are therefore likely to be different to the causes of ethnic residential segregation or occupational segregation by gender.

The paper is structured as follows. In the following section I outline a model of workplace segregation and show how preference spillovers lead to discontinuities in native workforce growth. In Section 3 I present the empirical implications of the model and discuss different tests for the existence of tipping points, before presenting the data to be used in Section 4. In Section 5 I present the results of tests for the presence of tipping points in the composition of both firms and local industries. Section 6 concludes.

2 Theoretical framework

2.1 A model of tipping

In this section I briefly adapt the model of Card et al. (2008, 2011) of neighbourhood composition in the presence of social interactions to segregation in the labour market. This stylised model will serve to structure the empirical analysis. The model is static and partial equilibrium. A representative, nondiscriminating firm hires two types of workers, immigrants and natives, denoted $j \in \{I, N\}$, which it treats as perfectly substitutable in production. The firm's size is taken as given, so the total workforce is normalised to equal one. The inverse supply of type j is given by $\omega^j(n_j, s)$, a primitive of the model. Crucially, the inverse supply depends not only on the quantity of workers of type j hired, n_j, but also on the share of immigrants in the firm, s.

The partial derivatives $\partial\omega^j(n_j, s)/\partial n_j$ are assumed to be weakly positive, that is, for a constant immigrant share, the firm needs to raise wages to hire more workers of a given type. The partial derivative $\partial\omega^j(n_j, s)/\partial s$ represents the social interaction effects. In particular, similar to Card et al. (2008), I assume that $\partial\omega^N(n_N, s)/\partial s > 0$ for s greater than some threshold;[5] that is, when the immigrant share in the firm is large, the firm needs to pay a higher wage to hire a given quantity of natives.

The assumption that $\partial\omega^N(n_N, s)/\partial s > 0$ for s sufficiently large is central to the derivation of tipping points. This assumption is consistent with heterogeneous underlying individual preferences. If all natives dislike working with immigrants, then $\partial\omega^N(n_N, s)/\partial s > 0$ for all s. If, consistent with survey evidence (European Commission, 2018), some natives are indifferent, or even positively inclined towards working with low levels of immigrants, then it may be the case that $\partial\omega^N(n_N, s)/\partial s \leq 0$ for low values of s. The only constraint on the underlying pattern of heterogeneity in native preferences is that the number of natives who for a given wage would prefer to take their outside option rather than work at the firm is increasing in s for s sufficiently large.

Under the normalisation that the total workforce is one, we have $n_N = 1 - s$, and the derivative of the native inverse supply function with respect to the migrant share will be

$$\frac{d\omega^N}{ds} = -\frac{\partial\omega^N}{\partial n_N} + \frac{\partial\omega^N}{\partial s}. \tag{1}$$

Under the previous assumptions, the first term will be negative, while the second term will be positive above some threshold. I follow Card et al. (2008) in assuming that the social interaction effect dominates at high levels of s, where $d\omega^N/ds > 0$, but not at low levels

[5]Note I am not assuming a discontinuity in $\partial\omega^N(n_N, s)/\partial s$; the partial derivative may vary smoothly through the threshold in s, or it may be positive for all $s \geq 0$.

of s. The supply of natives to the firm $n_N = 1 - s$ is therefore downward sloping in the wage for low levels of n_N; the reduction in s entailed by the increase in n_N increases the attractiveness of the firm sufficiently to attract more native workers, even at a lower wage. The native labour supply only becomes upward-sloping as n_N rises and the immigrant share s falls below a certain threshold. I also assume for simplicity that $d\omega^I/ds > 0$ for all $s \in (0, 1)$, that is, that supply of immigrants is upward-sloping in the wage for all values of n_I.[6]

There are multiple ways one could interpret the social interaction effects captured by the assumption that $\partial\omega^N(n_N, s)/\partial s > 0$. The simplest way, consistent with the original model of Card et al. (2008) and the tradition of social interactions models going back to Schelling (1971), is to interpret this as a consumption externality. Natives experience disutility from working with immigrants, so the marginal native worker will become unwilling to work at the firm if the immigrant share increases.

The source of this disutility could be a simple distaste or discomfort experienced by individual natives when working with immigrants. Alternatively, the disutility could arise from dynamic considerations, if natives believe that working with immigrants will harm their future job-finding prospects and earnings. Such beliefs could arise if immigrants are not a good source of referrals or information about job openings, or if an inflow of immigrants into a firm is a signal that the firm has experienced a negative productivity shock, as in the pollution model of Goldin (2014b).[7]

At an integrated equilibrium, where both types of workers are employed at the firm, the wages paid to both types of workers must be equal, since the firm is assumed to be non-discriminating. Again under the normalisation that the total workforce is one, an integrated equilibrium therefore requires that

$$\omega^N(1 - s, s) = \omega^I(s, s). \tag{2}$$

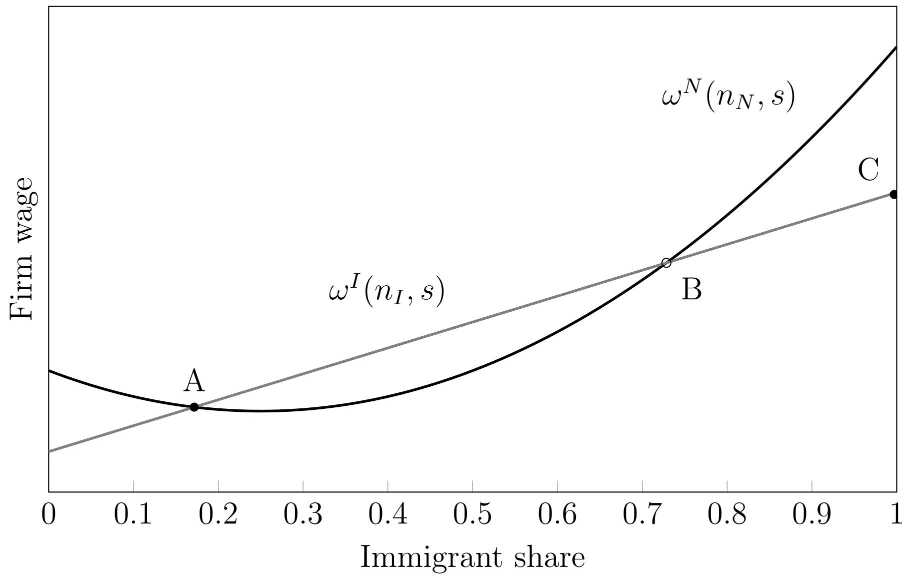
The inverse supply curves of immigrants and natives are plotted in Figure 2. As $s =$

[6] There is therefore an asymmetry in the strength of the social interaction effects between immigrants and natives that drives an asymmetry in the shape of the inverse supply curves of migrants and natives. This asymmetry is also present in the model of neighbourhood composition of Card et al. (2008). The empirical predictions of the model can still be derived when social interactions cause immigrant inverse supply to be downward sloping in n_I for low values of s; what is strictly necessary however is that the inverse supply curve of immigrants be flatter than the inverse supply curve of natives, i.e. $d^2\omega^I/ds^2 < d^2\omega^N/ds^2$, for all $s \in (0, 1)$.

[7] Alternatively, one could interpret the social interaction effect as a productivity externality, reinterpreting n_N as the effective supply of natives. Under this interpretation, an increase in the immigrant share lowers the productivity of natives; to keep a constant effective supply of native workers, the firm must raise the wage offered to hire more natives. This interpretation is consistent with recent evidence on negative productivity spillovers between immigrants and natives in certain firms (Glover et al., 2017). However, productivity spillovers would complicate the derivation of Equation (1), since now $n_I \neq s$. Furthermore, the empirical implications of the model would be unchanged, so I do not entertain this idea further here.

$n_I = 1 - n_N$, the supply of immigrants increases moving to the right on the x-axis, while the supply of natives increases moving to the left on the x-axis. As the inverse supply curves are drawn, there are two integrated equilibria (A and B) and one fully segregated equilibrium (C). Equilibrium A is stable in the sense that a small increase in the firm's minority share raises the wage that must be paid to immigrants above the wage paid to natives, so the firm hires natives until it returns to the equilibrium at A. The same remark holds *mutatis mutandis* for a decrease in the minority share at A or at C. Equilibrium B is, however, unstable. After a small increase in the immigrant share from B, the wage demanded by natives is greater than the wage demanded by immigrants, the firm will replace natives with immigrants until it reaches the equilibrium at C.

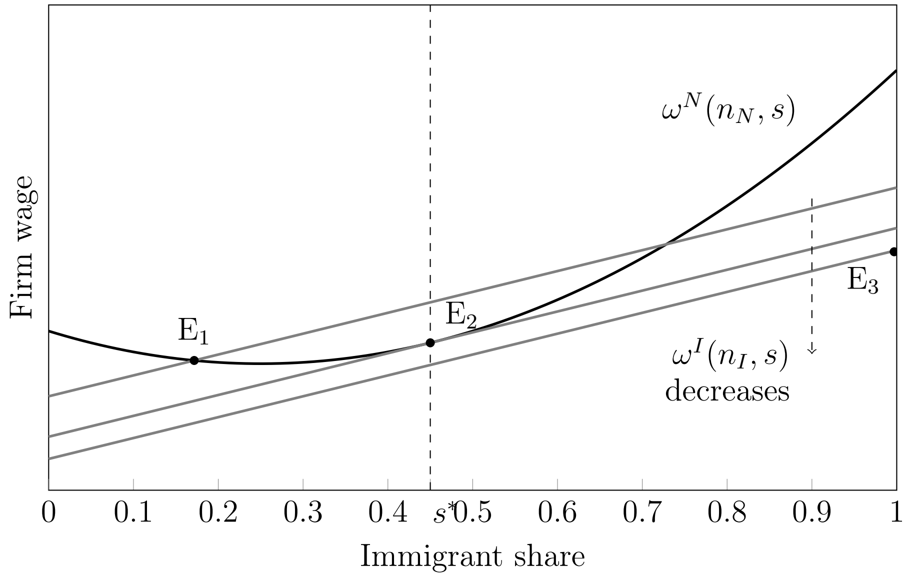
Figure 2: Immigrant and native inverse labour supply



Notes: Immigrant and Native inverse labour supply to the firm with three equilibria. A and C are stable, B is unstable.

In Figure 3 I plot what happens as the supply of immigrant workers to the firm increases exogenously, say, as a result of an inflow of immigrants to the local labour market where the firm is located. Suppose the firm is initially in equilibrium at E_1. An exogenous increase in the supply of immigrants shifts the immigrant inverse supply curve downward. The equilibrium moves to the right, eventually reaching the point of tangency E_2, which is stable with respect to decreases in the immigrant share, but unstable with respect to increases. If there are any further increases in the supply of immigrant workers, no integrated equilibrium will exist, the only equilibrium will involve the firm hiring only immigrants, as at point E_3. Traditional social interaction models such as Schelling (1971, 1978) or Becker and Murphy (2000) identify the unstable equilibrium B in Figure 2 as a tipping point. Here, however, I follow Card et al. (2008) in defining the tipping point as the maximum possible immigrant share in an integrated equilibrium. In Figure 3, this is

Figure 3: Effect of increasing supply of immigrant labour



Notes: Increasing supply of immigrant workers shifts their relative supply outwards, decreasing the wage demanded for any value of s. The equilibrium immigrant share starts at E_1 and shifts right as the inverse supply of migrants increases. The equilibrium E_2 is the maximum integrated equilibrium, the associated migrant share is s^*. If the supply of immigrant workers increases further, the firm will jump to the segregated equilibrium E_3, hiring only immigrants.

the immigrant share s^*, associated with the equilibrium E_2.

Two caveats are worth noting with this model. First, it does not account for the distribution of immigrants across firms, only the composition of a single firm. I implicitly assume that the natives who leave the firm after the tipping point is exceeded would either prefer to be unemployed than keep working in a high-immigrant-share firm, or are able to find jobs in other firms that have not faced a similar supply shock. Second, social interaction models are typically thought to lead to an inefficiently high degree of segregation across neighbourhoods, because agents cannot coordinate on where to locate. The model presented here, by only considering a single representative firm, is silent about the potential welfare consequences of such social preferences. It has traditionally been argued that firms, by internalising any spillovers across workers arising from their hiring decisions, choose a socially optimal degree of segregation (Becker and Murphy, 2000). However, these arguments do not account for the possibility that workplace segregation could be dynamically inefficient, if, for example, it keeps immigrants from developing the network or the kind of experience necessary to move up the job ladder.

2.2 Dynamic implications

While the model presented in the previous section is static, it is still possible to use it to make dynamic predictions about the composition of the representative firm's workforce.

Consider a firm whose initial static equilibrium immigrant share is $\bar{s}_0 < s^*$, where s^* is the tipping point defined previously as the immigrant share associated with the maximum possible integrated equilibrium. Suppose the firm experiences a small increase in the supply of immigrants, i.e. a fall in the wage a given quantity of immigrant labour needs to be paid, $\Delta\omega^I(n_I, s) < 0$, between period 0 and period 1.[8] There will be some $r \in (0, s^*)$ such that if $\bar{s}_0 \in [0, s^* - r)$, the firm's new equilibrium will be at $\bar{s}_1 \in (0, s^*]$, whereas if $\bar{s}_0 \in [s^* - r, s^*]$, the increase in the immigrant supply takes the firm beyond the point of tangency at E_2 in Figure 3 and the new equilibrium will be $\bar{s}_1 = 1$. As the increase in the immigrant supply $\Delta\omega^I(n_I, s)$ becomes infinitesimally small, r also approaches zero. Note that no firm can initially be at an equilibrium at $\bar{s}_0 \in (s^*, 1]$ except for at $\bar{s}_0 = 1$, where a small increase in the supply of immigrants will have no effect on the equilibrium.

Assume that the firm myopically adjusts its immigrant share in response to changes in the supply of immigrants such that the immigrant share s_t remains close to its equilibrium value. To allow for the possibility that search or other labour market frictions prevent the immigrant share from fully adjusting within a single period to a new equilibrium value as the supply of immigrants changes, I use the notation s_t to refer to the observed immigrant share at a point in time, to distinguish it from the static equilibrium at that point in time, $\bar{s}_t$. For an observed $s_0 \in [0, s^* - r)$, the observed increase in the immigrant share Δs_1 in response to the increase in the immigrant supply $\Delta\omega^I(n_I, s)$ will be small. However, for $s_0 \in [s^* - r, s^*]$, $\Delta\omega^I(n_I, s)$ will cause a large observed Δs_1, as the firm converges to the new equilibrium at $\bar{s}_1 = 1$. For firms initially at $s_0 \in (s^*, 1)$, the tipping process is already underway, and one should expect to see $\Delta s_1 > 0$ and larger the closer the firm is to s^*. There will therefore be a discontinuity in Δs_1 around the tipping point s^*. We will observe Δs_1 to be small and positive for s_0 to the left of the tipping point and large and positive for s_0 close to or beyond the tipping point.

Whilst the foregoing discussion restricts attention to the case of an increase in the immigrant supply, where the discontinuity appears clearly, the discontinuity will also exist in the case where there is a decrease in the immigrant supply. This is because once a firm has started tipping and $s_0 \in (s^*, 1]$, a small decrease in the supply of immigrants will typically not reverse the tipping process, implying that for these firms too $\Delta s_1 > 0$. The condition for tipping to continue after a decrease in the immigrant supply is for the marginal immigrant to continue to accept a lower wage than the marginal native, which is

[8] The discussion here in fact holds for an increase in the relative supply of immigrant, $\omega^N(n_N, s) - \omega^I(n_I, s)$. However, to simplify the discussion I assume the supply of natives is fixed and only the supply of immigrants varies.

more likely to be satisfied the smaller the decrease in the immigrant supply or the further to the right of s^* the firm initially finds itself. On the other hand, for a firm that is close to tipping, but where $s_0 < s^*$, a small decrease in the immigrant supply will lead to a small decrease in the immigrant share in the firm.

Combining these observations about the effect of increases and decreases in the immigrant supply on the firm's immigrant share, one can conclude that there will be a discontinuity in the expected change in the immigrant share as a function of the base-year immigrant share:

$$\mathrm{E}[\Delta s_t | s_{t-1}] = \mathbf{1}(s_{t-1} < s^*)g(s_{t-1}) + \mathbf{1}(s_{t-1} \geq s^*)h(s_{t-1}) \tag{3}$$

where $\lim_{\epsilon \to 0^+} h(s^* + \epsilon) - g(s^* - \epsilon) > 0$. $h(s_{t-1}) > 0$, while the sign of $g(s_{t-1})$ will depend on whether firms more commonly face increases or decreases in the immigrant supply. The existence of a discontinuity in $\mathrm{E}[\Delta s_t | s_{t-1}]$ at the tipping point s^*, which does not depend on whether the immigrant supply is increasing or decreasing, is the key dynamic implication of the model I will test in the empirical analysis below.

3 Empirical implementation

3.1 Unit of analysis

While the model presented above predicts that tipping points might be observed in the composition of a firm's workforce, one might also expect to observe tipping dynamics in the composition of larger aggregates, such as the industry, occupation, or geographic area. Indeed Goldin (2014b) notes that the pollution model she develops to explain the dynamics of workplace composition by gender might operate at the level of firms, occupations, industries, or geographic aggregates. Historically, there is evidence in France at least of high immigrant shares in an industry being associated with low prestige of the industry (Noiriel, 1988), suggesting that tipping might occur in the composition of larger aggregates. On the other hand, if the kinds of preference spillovers underpinning the model of tipping presented above are experienced primarily in direct personal interactions in the workplace, as in the cases studied by Hjort (2014) or Glover et al. (2017), one might expect to only observe tipping in the composition of firms.[9]

Empirically, segregation across firms and across larger units of aggregation appear to be distinct phenomena. Table 1 reports the index of coworker segregation—defined

[9]Note also that there is no straightforward logical relationship between tipping at, say, the industry level and at the firm level. Industry-level tipping does not imply firm-level tipping, since it could occur through the entry of high immigrant-share or the exit of high native-share firms as the industry passes the tipping point. Similarly, tipping at the firm level might only imply a reallocation of a fixed pool of workers within the industry, leaving the aggregate composition unchanged.

by Hellerstein and Neumark (2008) as the excess probability that an immigrant has of working with other immigrants, relative to a native—for West Germany in 1990–2010. Throughout this period, an immigrant was at least 15 percentage points more likely to work with another immigrant than natives were. The index of coworker segregation can be normalised to account for differences in the distribution of immigrants and natives across larger units of aggregation, such as regions or industries, yielding what is known as an *effective* index of coworker segregation. Conditioning the index on the distribution of workers over local labour markets and three-digit industries reduces an immigrant's excess probability of working with other immigrants to 8–10 percentage points, explaining around 45 per cent of observed segregation, with segregation across industries appearing to contribute more to this reduction than segregation across locations.

Table 1: Index of coworker segregation

	1985 ICS	1990 ICS	1995 ICS	2000 ICS	2005 ICS
Unconditional	0.15	0.15	0.17	0.18	0.17
Conditional on industry	0.12	0.12	0.13	0.14	0.13
Conditional on location	0.14	0.14	0.16	0.17	0.16
Conditional on location and industry	0.08	0.08	0.09	0.10	0.10
Establishments	480174	520285	545225	721179	745427

Note: Indexes of coworker segregation of Hellerstein and Neumark (2008), calculated from the *Betriebshistorikpanel* of the IAB. Includes all establishments in West Germany hiring two or more employees. The conditional indexes condition on either three-digit industry (NACE Rev. 1), local labour market, or both.

3.2 Identifying the location of the tipping point

Any test for the existence of a tipping point in workforce composition needs to reckon with the fact that the theoretical tipping point s^* is unknown. Card et al. (2008) propose treating identifying the location of the tipping point and testing for the existence of a tipping point as separate problems and solving them sequentially. In the first step, they use a search procedure to identify a candidate tipping point. The simplest procedure they propose is a threshold regression (Hansen, 2011, 2021). In the second step, Card et al. (2008) use regression discontinuity design (RDD) techniques (Imbens and Lemieux, 2008; Lee and Lemieux, 2010) to estimate Equation (3). If the estimated discontinuity in the change in the minority share when the minority share moves beyond the candidate tipping point is negative and significant, they conclude that there is a tipping point in the composition of the units under study.

To address the possibility of specification search bias that would arise when using the same data to both identify the location of the tipping point and estimate the size of the discontinuity at the tipping point, Card et al. (2008) propose two solutions. Either the researcher can split the sample, using independent subsamples for the search and estimation steps described above, or the researcher may bootstrap the entire two-step procedure to construct standard errors for the estimated second-stage discontinuity.

The estimation and inference procedure proposed by Card et al. (2008) has been adopted, essentially unmodified, in most subsequent tests of tipping points (Aldén et al., 2015; Böhlmark and Willén, 2020; Pan, 2015; Zheng, 2014).[10] However, the approach suffers from two shortcomings. First, treating the second stage as an RDD is arguably conceptually incorrect, since there is no treatment variable whose assignment probability jumps at the threshold, other than the tautologically defined treatment "being above the tipping point". So while one may still use local polynomials to descriptively estimate a break in the outcome variable at the candidate tipping point, the standard RDD interpretation of this break as an average treatment effect does not apply.[11] Second, the inference procedures proposed by Card et al. (2008) may not be suitable in all settings, and in particular in settings where there is in fact no tipping point.

Conducting inference via sample splitting is not efficient, since only a subset of the data is used at either stage. The approach therefore relies on the availability of a large dataset, which is the case when studying tipping in firms or neighbourhoods, of which there are many, but not when studying larger aggregates, such as local industries. Conducting inference via the bootstrap, on the other hand, is feasible in smaller datasets, but its validity has been demonstrated under the assumption that the discontinuity being estimated is large relative to sampling variation (Hansen, 1996; Elliott and Müller, 2007). However, in situations where it is not obvious from simply looking at the data whether there is a tipping point or not, such an approach may lead to over-rejection of the null hypothesis of no tipping points (Andrews et al., 2021). In such settings, the researcher may conclude that there are tipping points where there are in fact none.

To address these shortcomings, I propose to test for the presence of tipping points by estimating a single threshold regression and using inference procedures that are robust to small effects. Both the location of the tipping point and the size of the break in the outcome are estimated via a threshold regression that takes the following general form:

$$Y_{it} = C_{it}'\beta + D_{it}'\delta\mathbf{1}\{Q_{it} > \theta\} + u_{it}. \tag{4}$$

Let the number of immigrants employed in firm i at time t be I_{it}, the number of natives

[10] An exception is Caetano and Maheshri (2017), who propose an alternative method.

[11] This criticism does not apply to the work of Böhlmark and Willén (2020), since for their main results they use tipping points identified via a Card-style procedure as discontinuities in an unrelated RDD, where the main outcome is individual-level educational attainment.

be N_{it}, and the total workforce $L_{it} = I_{it} + N_{it}$. Following Pan (2015), the dependent variable Y_{it} is defined as the five-year change in the native workforce, normalised by the base year workforce, minus the normalised five-year change in the immigrant workforce: $Y_{it} = (N_{it+5} - N_{it})/L_{it} - (I_{it+5} - I_{it})/L_{it}$.[12] The change in immigrant demand is therefore a proxy for changes in total workforce demand, which are netted out in this formulation (Pan, 2015). The vector of control variables C_{it} includes a polynomial function in the base-year immigrant share and other base-year controls depending on whether the unit of observation is the firm or industry. Q_{it} is the base year immigrant share and θ is the tipping point. The set of variables D_{it} is the subset of C_{it} whose effect on Y_{it} varies when the base-year immigrant share passes the tipping point. In my specifications D_{it} only includes a constant; in this case, the parameter δ measures the key discontinuity. We conclude that there is a tipping point if δ is negative and significant. The estimation Equation (4) is the empirical counterpart of Equation (3).

Equation (4) is nonlinear in the parameter vector $(\beta', \delta', \theta)'$, and is estimated by non-linear least squares (NLS). The location of the tipping point, θ, and the size of the discontinuity at the tipping point, δ, are therefore estimated simultaneously. The difference between this approach and that of Card et al. (2008) bears emphasising. They estimate the location of the candidate tipping point s^* from a simple threshold regression where $C_{it} = D_{it} = \iota$, a constant, and then estimate $\delta(s^*)$ from a follow-up OLS regression of Equation (4) where they set $\theta = s^*$ and C_{it} includes higher-order polynomial terms and other controls.

While Equation (4) can be estimated by NLS, the parameters are not asymptotically normally distributed, since θ is not identified when $\delta = 0$ (Hansen, 2021). Hansen (1996) has shown that a bootstrap procedure will yield correct p-values for the test that $\delta = 0$, and Card et al. (2008) appeal to this result when justifying the use of the bootstrap to construct standard errors for $\delta(s^*)$ in their two-step procedure. However, the validity of the bootstrap procedure in the threshold regression setting is shown under a restrictive set of assumptions (Andrews et al., 2021). As a result, Andrews et al. (2021) propose an alternative procedure for constructing standard errors for δ when estimating a threshold regression. In particular, their procedure is robust to (i) the true threshold effect δ being small relative to sampling variation; and (ii) the model (4) being misspecified, which is likely if Equation (4) is only a parsimonious approximation of the true conditional expectation of Y_{it}. I will therefore use the so-called "hybrid" standard errors proposed by Andrews et al. (2019, 2021) when conducting inference on δ. These standard errors have

[12] Other tests of tipping points, such as Card et al. (2008) or Pan (2015), study ten-year changes. In part this is due to data constraints, since those papers use decennial census data. Since the costs of changing workplace are arguably lower than changing residence, workplace tipping dynamics will appear on a shorter time scale. Focusing on ten-year changes would also lead to greater selection into the sample via firm exit, which is potentially correlated with the base-year immigrant share.

been shown both theoretically and in simulations to have good coverage properties both when the truth is $\delta = 0$ and when $\delta \neq 0$. The interested reader is referred to Andrews et al. (2019, 2021) for full details on the construction of these standard errors.[13]

3.3 Details

The location of the true tipping point, $\theta = s^*$, will depend on the shape of the inverse supply curves of immigrant and native workers. Various factors will affect the shape of the inverse supply functions, two of which bear emphasising here. The first factor is heterogeneity across labour markets in individual tastes within the pool of workers the firm might potentially hire, and, in particular, heterogeneity in the strength of native distaste for immigrants, measured by the partial derivative $\partial \omega^j(n_j, s)/\partial s$. If the value of the partial derivatives of the inverse supply functions is the same across labour markets, then the tipping point will also be the same for different labour markets. Both Card et al. (2008) and Aldén et al. (2015) assume different tipping points for different residential markets, while Pan (2015) assumes the location of tipping points in labour markets varies by region-occupation type (white/blue collar) cell.

The strength of native distaste for immigrants likely varies with the level of historical exposure to immigrants, which varies across locations, and possibly also across industries. In my main specification I will therefore follow the previous literature and assume the location of the tipping point varies by region (German *Bundesland*, where smaller states are merged with larger neighbours) and industry type (agriculture, manufacturing, mining, construction, and hospitality versus other industries). I therefore estimate Equation (4) separately for each region by industry type and for each base year, since tipping dynamics might be observed in some years but not others. There are 14 such region-by-industry cells in my data in each year.

The second important factor that will affect the location of the tipping point are firm-specific amenities differentially valued by natives and immigrants. Such amenities may make some firms more attractive to natives than others, for a given wage and immigrant share, potentially altering the shape of the native inverse supply curve. If such amenities vary substantially across firms, the location of the tipping point will also vary across firms.[14]

[13] Andrews et al. (2021) develop their procedure in the case where $D_i = C_i$. Since in my setting D_i is typically a constant while C_i includes a polynomial in the base-year immigrant share, and since the method of Andrews et al. (2021) has not yet to my knowledge been used in applications, I present in detail the changes that are necessary to implement their method and construct standard errors when D_i is a subset of C_i in Appendix B, available online. These details may be of use to researchers interested in implementing the procedure of Andrews et al. (2021) in other settings.

[14] The role of heterogeneous neighbourhood amenities in the dynamics of residential segregation has been highlighted theoretically by Banzhaf and Walsh (2013) while Caetano and Maheshri (2017) propose a method for testing for the presence of school-specific tipping points in school composition given hetero-

The procedure presented in Section 3.2 assumes there is a tipping point that is common to at least some subset of firms, i.e. that it is possible to group firms by non-wage amenities before testing for a common tipping point among these firms. For this assumption to hold, a good proxy for firm-level amenities must be used to group firms. Sorkin (2018) has found that 70 per cent of the variance of the firm component of wages, estimated as firm fixed effects, reflect compensating differentials for firm-level amenities. Furthermore, 30–45 per cent of the variation in compensating differentials is explained by industry, which suggests that most firms within sufficiently small industry cells should share a tipping point, if the tipping point exists. In Section 5.3, I confirm that my results hold when I assume that the tipping point is common to (i) firms with similar estimated wage fixed effects; or (ii) firms that belong to the same three-digit industry.

4 Data

The data used to test for the presence of tipping points in the German labour market come from the Institute for Employment Research of the German Federal Employment Agency (IAB). I use the Establishment History Panel (BHP), a fifty per cent sample of all establishments making social security contributions for at least one employee between 1975 and 2019.[15] An establishment covers all production sites belonging to the same firm, located within the same municipality, and operating within the same three-digit sector. I follow standard practice when working with the BHP in indiscriminately referring to establishments as firms or establishments.

The sampling frame of the BHP includes all firms making social security contributions in West Germany since 1975, and all such firms in East Germany since 1993. I restrict my attention to the period 1975–2010 and separately analyse changes over each of the seven five-year periods in the dataset, starting from 1975–1980. This allows me to investigate potential differences in tipping dynamics as immigrant flows and macroeconomic conditions change over time. I also limit myself to West Germany (excluding Berlin) since East Germany is not covered through the whole period and a large majority of Germany's immigrants live and work in the old West Germany.

I test for tipping dynamics in both the composition of firms and of local industries. When studying firms, I impose the supplementary restriction that firms employ at least 10 workers in the base year. I do this since (i) the immigrant share variable is not continuous when there are few employees and has mass points around values such as 0.25, 0.33, or 0.5, while the theory developed in Section 2 assumes the immigrant share

geneous school amenities. Extending their approach to the case of firms, which would require a credible instrument for the immigrant share in the firm, is beyond the scope of the present work.

[15]Specifically, I use version 2 of the 1975–2019 edition of the BHP. For details on this dataset, see Ganzer et al. (2021).

is continuously distributed; (ii) the immigrant share can change dramatically over time when there are only a few workers, creating artificial discontinuities in Y_{it} around the values of the base year immigrant share where there are mass points; and (iii) small firms are more likely to enter or exit over a five-year period, potentially creating sample selection issues. When studying local industries, I similarly impose the restriction that the industry be constituted of at least ten firms, employing at least 30 workers between them. I further exclude both industries and firms where either the normalised native or immigrant workforce growth exceeds 300 per cent over five years, since the theory in Section 2 assumes the firms size is constant.

Table 2: Summary statistics

	(1) 1975	(2) 1980	(3) 1985	(4) 1990	(5) 1995	(6) 2000	(7) 2005
A: Aggregate Statistics							
Region-industry cells	14	14	14	14	14	14	14
Immigrant share	9.5	9.1	7.3	7.8	9.4	8.3	8.0
Employment growth	4.2	-1.9	8.4	-3.4	16.3	-0.2	5.8
Native growth	4.3	-0.06	7.3	-4.5	16.0	0.3	5.2
Immigrant growth	-0.05	-1.9	1.1	1.1	0.3	-0.5	0.6
B: Firm Statistics							
Share of employment	64.5	64.2	64.2	64.7	63.0	62.3	63.3
Immigrant share	7.1	6.6	5.6	6.2	8.1	6.6	6.6
Employment growth	4.8	-4.1	7.6	1.2	15.7	-1.3	5.1
Native growth	4.7	-2.9	6.4	-1.0	16.0	-1.0	4.4
Immigrant growth	0.1	-1.2	1.2	2.2	-0.4	-0.3	0.7
Firms	96360	106319	105199	115476	117884	158135	167374
C: Industry Statistics							
Share of employment	84.0	85.1	85.7	86.8	88.2	91.2	91.5
Immigrant share	7.0	6.6	5.3	6.1	8.0	6.5	6.3
Employment growth	15.8	4.3	17.8	8.3	46.9	6.7	11.1
Native growth	15.2	5.3	15.9	5.7	44.8	6.3	9.8
Immigrant growth	0.6	-1.0	1.9	2.6	2.0	0.4	1.3
Local industries	5725	6093	6308	6746	6802	7894	7935

Note: Panel A reports aggregate statistics for all of West Germany using the BHP of the IAB. Panel B reports averages for the included firms; Panel C reports averages for the included local industries (three-digit industries by local labour markets). Growth rates are expressed in percentage terms for the five-year period starting in the base year defined for each column. Immigrant growth and native growth are normalised by total base-year employment.

Aggregate summary statistics, using all BHP firms in West Germany, are presented in Panel A of Table 2. Averages over the firms included in my sample are in Panel B, while averages over the included local industries are in Panel C. The size restrictions imposed mean that the sample of firms cover around 62–65 per cent of total employment subject to social security in West Germany, while the sample of local industries covers around 84–92 per cent of employment. The average immigrant share in the firms and local industries studied is a little lower than in the full set of firms, implying, via Bayes's rule, that the sample of firms covers around 50 per cent of employed immigrants in Germany, while the sample of local industries covers around 65 per cent of employed immigrants.

At both levels of observation, the average immigrant share falls during 1975–1985, increases in 1985–1995, and falls again somewhat thereafter, mirroring net migration flows to Germany over the time period. Given that tipping dynamics are observed when there is an increase in the relative supply of immigrants facing a firm or local industry, this suggests that tipping dynamics are more likely to be observed in the period 1985–1995. While average normalised native and immigrant workforce growth are broadly correlated, there are periods, in particular 1990–1995 when high immigrant inflows, in this case linked to wars in ex-Yugoslavia, coincided with protracted recessions, leading immigrant employment to grow on average even as total employment contracted.[16]

5 Results

5.1 Firms

To test for the presence of tipping points in firms, I estimate Equation (4) separately for region-by-industry type cells. The dependent variable is modelled as a fourth-order polynomial in the base-year immigrant share, with an intercept shift at the tipping point, including the log of the median wage of a native in the firm, the low-skilled workforce share, and the firm's share of total employment in the local industry as additional controls. In the first section of Table 3, I report the average estimated break point in the threshold model, which ranges between base year immigrant shares of 36 and 48 per cent. The average NLS estimate across cells of the discontinuity ranges from -9 to 24, and is negative during the period 1975–1985 and positive thereafter. The largest negative average discontinuities are for the periods 1980–1985 and the largest positive ones are in 2000–2005. The fact that the average discontinuity in the period with the largest immigrant inflows, 1990–1995, and in the period after, 1995–2000, is small and positive should already invite scepticism about the presence of tipping points in the composition of firms in Germany.

I next report the share of cells in which the estimated discontinuity is significant,

[16] Germans migrating from the old East Germany to the West are classified as Germans, not migrants.

Table 3: Estimated discontinuities in firm net native workforce growth

	(1) 1975	(2) 1980	(3) 1985	(4) 1990	(5) 1995	(6) 2000	(7) 2005
Tipping point	36.1	40.8	45.5	48.3	42.4	39.7	48.2
	(26.6)	(31.9)	(37.0)	(34.7)	(31.7)	(32.7)	(31.1)
Discontinuity ($\hat{\delta}$)	-4.4	-8.9	4.9	2.4	1.8	23.6	12.5
	(33.1)	(37.4)	(51.0)	(61.0)	(43.6)	(72.4)	(32.5)
Median-unbiased $\hat{\delta}$	-5.0	-6.3	8.3	3.6	3.4	18.4	10.2
	(31.9)	(31.3)	(36.2)	(56.7)	(36.7)	(73.0)	(30.1)
Share p-val. < 0.05	0.43	0.14	0.14	0.36	0.07	0.14	0.57
p-val. < 0.05 and $\hat{\delta} < 0$	0.29	0	0.07	0.14	0	0	0.50
Median LB, 95% CI	-19.8	-12.1	-19.3	-10.3	-12.6	-12.8	-15.0
Median UB, 95% CI	6.4	11.2	23.1	15.1	16.4	17.5	-0.04
Region-industry cells	14	14	14	14	14	14	14
Avg. obs.	6867	7576	7496	8229	8392	11036	11536

Note: The first section reports averages and standard deviations (in parentheses) of the estimated discontinuity $\hat{\delta}$ across region by industry-type cells. The second section reports measures of average significance, the third section reports the sample sizes. See text for details on estimation and inference procedures used as well as sample definitions.

using the so-called hybrid standard errors proposed by Andrews et al. (2019, 2021).[17] The estimated discontinuity is significant in at most 57 per cent of region-by-industry type cells (eight out of 14) and is often significant in only one or two out of 14 cells. Furthermore, the estimated discontinuity is negative and significant in only one out of 14 cells in 1985–1990, two out of 14 cells in 1990–1995, and in no cells at all in a further three periods. There therefore appears to be very little evidence for the presence of tipping points in the composition of firms, across the sample.

A lack of evidence for the presence of tipping points is not the same as evidence against there being tipping points in the composition of firms. To establish that there likely are no tipping points in the composition of firms, first note that the average sample sizes in each year are relatively large, ranging from around 7,000 firms to 11,000 firms across periods, so precision is not likely to be a problem. To further understand how precisely the discontinuity is estimated, I report the median lower bound for 95 per cent confidence interval for the estimated discontinuity across cells within a given year, and the median upper bound. The median lower bound of a 95 per cent confidence interval is on average

[17] The term hybrid refers to the fact that these standard errors combine the desirable features of (i) standard errors that have correct coverage rates conditional on the true breakpoint being equal to the estimated breakpoint, but that can have infinite expected length; and (ii) standard errors that have constrained length but only have correct coverage properties on average for all possible breakpoints, and not necessarily for the specific estimated breakpoint.

-15 percentage points. The median upper bound averages 13 percentage points, although 2005 is an exception, where the median is close to zero.

Comparing the size of these confidence intervals with previous findings of tipping points, we can conclude that being able to repeatedly rule out negative discontinuities larger than 15 percentage points in magnitude with 95 per cent confidence should be considered strong evidence against the presence of tipping points. In the periods where tipping in the gender composition of occupations is strongest, Pan (2015) finds average discontinuities in net male employment growth of around 60 percentage points, albeit using a slightly different specification. In white-collar occupations where the evidence is strongest, the estimated discontinuities even rise to over 80 percentage points. Card et al. (2008) find smaller average discontinuities in neighbourhood composition, around 15–20 percentage points in comparable cell-by-cell regressions (what they label their "fully interacted" specification). The smaller magnitude Card et al. (2008) report is consistent with moving costs being higher for neighbourhoods than for occupations.[18]

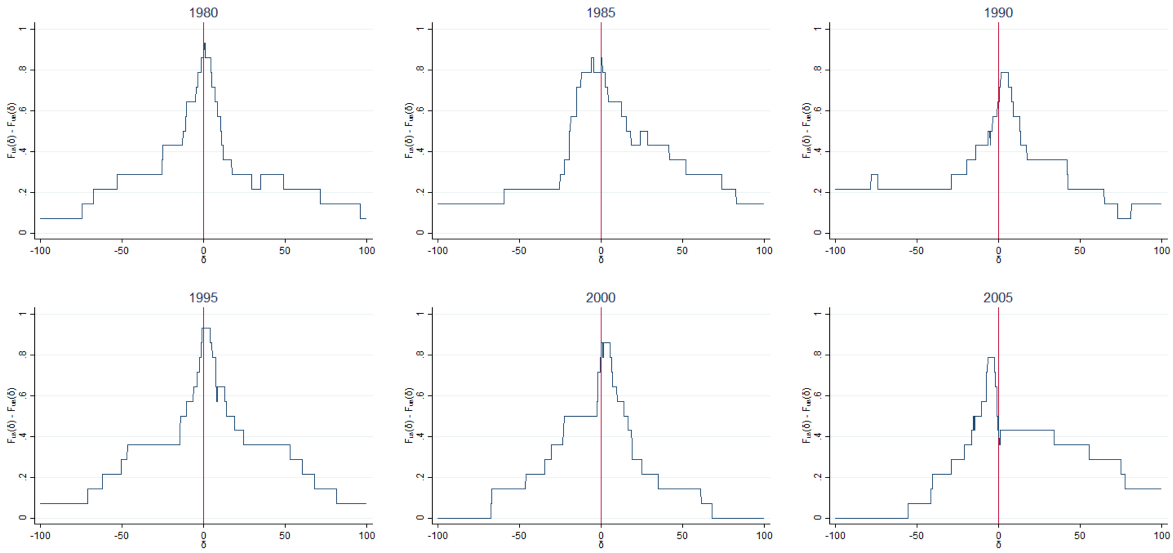
An alternative approach to evaluating the strength of the evidence against tipping points in the composition of firms is to calculate empirical coverage probabilities for a range of possible discontinuities. These are presented in Figure 4. For each base year and each possible value of the discontinuity $\delta \in [-100, 100]$, I calculate the share of confidence intervals that include the proposed value of the discontinuity δ. Since the confidence intervals are constructed using the hybrid standard errors of Andrews et al. (2019, 2021), they have been shown theoretically and in simulations to have correct coverage rates. If there are in fact no discontinuities, the empirical coverage rate at zero should have a value of approximately 0.95 for 95 per cent confidence intervals.

This expected pattern is what I effectively observe in practice. The empirical coverage rate at zero is lowest in 2005, where it is around 0.5, for no other year is the empirical coverage rate at zero below 0.8. More importantly, the coverage rates are always maximised at or within a couple of percentage points of zero, and typically decrease rapidly as the value of the discontinuity moves away from zero. Note that the empirical coverage rate at the true value may depart from the expected 0.95 coverage rate because it is calculated from relatively few regressions, 14 for each base year, so the law of large numbers does not apply. This explains why we do not observe coverage rates of exactly 0.95 for any particular value.[19] All in all, the evidence from the empirical coverage rates

[18]These two papers report discontinuities for ten-year changes. While one might expect the estimated discontinuity to be somewhat smaller for five-year changes, job switching costs are likely lower than occupation switching costs, and certainly lower than neighbourhood switching costs. Dustmann et al. (2016) report that median job tenure in Germany is 2–3 years, suggesting that five years is a sufficiently large window to observe any changes in workforce composition.

[19]Coverage may also be less than the nominal 0.95 rate if the true value of the discontinuity is in fact different across cells. However, given that there is little evidence of *any* discontinuities in net native workforce growth, there is little empirical basis to assume that the magnitude of the true discontinuity varies across cells.

Figure 4: Empirical coverage rates, firms



Notes: The figures plot the empirical coverage rates for different values of $\delta \in [-100, 100]$. The empirical coverage rate is defined as the empirical CDF of the lower bound of the 95 per cent confidence interval for $\hat{\delta}$ minus the empirical CDF of the upper bound of the 95 per cent confidence interval for $\hat{\delta}$.

strongly supports the conclusion that there are no tipping points in the composition of firms' workforces.

5.2 Industries

As noted previously, the absence of tipping points in the composition of firms does not rule out the possibility that these might be present in the composition of industries. To test for the presence of tipping points in the composition of industries, I estimate Equation (4) over local industries, using the same fourth-order polynomial specification with an intercept shift and including controls for log median native wage in the industry, share of low-skilled employment, average firm size, and the Herfindahl-Hirschman index of employment concentration in the local industry. The regressions are again run separately for each region-by-industry type cell. Table 4 first presents the average estimated tipping point, which typically corresponds to a base-year immigrant share between 10 and 16 per cent. The average NLS estimate of the discontinuity in net native employment growth is positive in most years, with the exception of 1980 and 1995, while the average estimate of the discontinuity using the median-unbiased estimator for the estimation of breaks does not differ substantially from the NLS estimates.

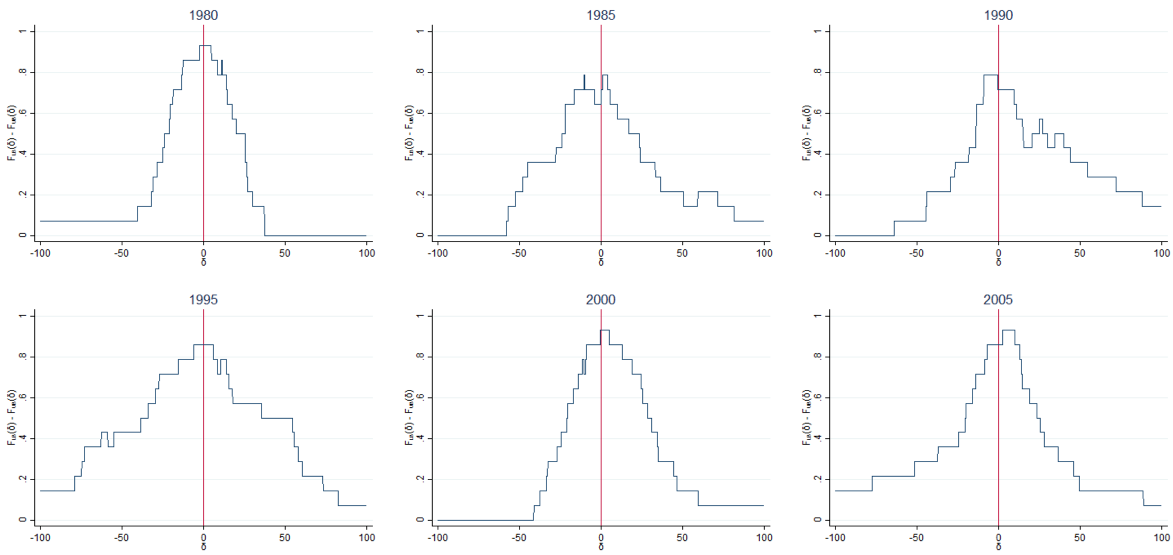
As with the firm estimates, the estimated discontinuities in net native workforce growth across industries is typically only significant in a handful of cells out of 14, and is not significant and negative in more than two out of 14 cells for any base year. The estimates

Table 4: Estimated discontinuities in industry net native workforce growth

	(1) 1975	(2) 1980	(3) 1985	(4) 1990	(5) 1995	(6) 2000	(7) 2005
Tipping point	12.2	10.0	11.5	13.5	16.3	13.2	14.9
	(5.7)	(4.8)	(5.0)	(6.0)	(11.6)	(8.5)	(10.8)
Discontinuity ($\hat{\delta}$)	28.5	-6.6	3.8	25.7	-6.2	12.3	14.3
	(68.1)	(23.9)	(36.1)	(41.3)	(53.1)	(32.5)	(60.5)
Median-unbiased $\hat{\delta}$	21.5	-4.8	4.0	24.5	-16.8	8.1	15.2
	(69.2)	(22.3)	(34.3)	(40.1)	(48.6)	(31.4)	(57.9)
Share p-val. < 0.05	0.07	0.07	0.29	0.29	0.14	0.07	0.14
p-val. < 0.05 and $\hat{\delta} < 0$	0	0	0.14	0.07	0.07	0.07	0
Median LB, 95% CI	-23.0	-22.7	-23.2	-14.1	-47.0	-21.0	-20.5
Median UB, 95% CI	38.9	22.6	23.5	34.8	44.8	29.5	26.4
Region-industry cells	14	14	14	14	14	14	14
Avg. obs.	409	435	451	482	486	564	567

Note: The first section reports averages and standard deviations (in parentheses) of the estimated discontinuity $\hat{\delta}$ across region by industry-type cells. The second section reports measures of average significance, the third section reports the sample sizes. See text for details on estimation and inference procedures used as well as sample definitions.

Figure 5: Empirical coverage rates, industries



Notes: The figures plot the empirical coverage rates for different values of $\delta \in [-100, 100]$. The empirical coverage rate is defined as the empirical CDF of the lower bound of the 95 per cent confidence interval for $\hat{\delta}$ minus the empirical CDF of the upper bound of the 95 per cent confidence interval for $\hat{\delta}$.

for industries are, however, relatively less precise than for firms. The median lower bound ranges between -14 and -47 percentage points, while the median upper bound ranges between 23 and 45 percentage points. This lower precision is largely due to the fact that there are many fewer industries than there are firms. The average cell-specific regression is run on a sample of between 409 and 567 observations, an order of magnitude smaller than for the firm regressions.

Given the industry estimates are relatively less precise than the firm estimates, it is more difficult to conclude from the absence of evidence for the presence of tipping points that there are indeed no tipping points in the composition of industries. However, the empirical coverage rates of the nominal 95 per cent confidence interval, presented in Figure 5, can again provide support for the conclusion that there are no tipping points in the composition of local industries. The empirical coverage rates are again maximised at or close to zero, always taking maximum values not less than 0.8. The empirical coverage function decreases more slowly as the proposed value of the discontinuity moves away from zero, reflecting the smaller sample size. As in the case of firms, the empirical coverage probabilities provide relatively strong evidence against there being tipping points in the composition of industries, in spite of smaller sample sizes.

5.3 Robustness

Here I consider several alternative explanations for the lack of evidence of tipping points in the compositions of firms or industries. In the case of firms, one might contend that the correct level of analysis is in fact the production team, not the firm, since it is within such teams that the interpersonal interactions with immigrants in which natives may experience disutility take place. Tipping in the composition of production teams might lead to sorting across production teams within the firm, without necessarily leading to observable tipping dynamics in the overall composition of the firm.

Since I do not observe information on individual workers' occupations or on the composition of firms, I cannot directly test for tipping points at the sub-firm level. However, to establish that the firm is not too large a unit of analysis, I repeat my main estimation specification, limiting the sample to small and medium-sized firms, i.e. those firms employing 10–49 workers. I report the results of these specifications in Table A.1. The average location of the estimated breakpoint is similar to when using the full sample. The average negative discontinuity is now negative in five periods, however, it is only greater than -10 percentage points in 1980 and 1990. The share of cells for which the discontinuity is negative and significant is not larger than when using the full sample, never exceeding 0.21, or three cells out of 14. There therefore does not appear to be strong evidence in favour of the existence of tipping points in the composition of smaller firms.

Concluding from this evidence that there are no tipping points in the composition of small firms is a little more difficult since the estimates are less precise. The median lower bound ranges between approximately -10 and -44 for different periods, and the median upper bound ranges between approximately 8 and 60. The evidence from the empirical coverage probabilities, reported in Figure A.1 is also less conclusive than when considering all firms. The empirical coverage rate sometimes appears bimodal, albeit around two values that are close to zero, and the maximum empirical coverage rate is typically between 0.7 and 0.8 and close to zero. Taken together, the evidence from the coverage rates still points to there being a low probability of tipping dynamics in the composition of medium firms.

A crucial assumption when implementing a test for tipping dynamics is the set of firms or industries which share a common tipping point, here assumed to be firms in the same region and broadly defined industry. However, as mentioned in Section 3.3, if amenities vary too much within region-industry cells, firms will not share a common tipping point.

To address this possibility, I first divide firms into ventiles of the firm fixed effect from an individual wage regression.[20] Variation in firm wage fixed effects has been shown to be largely driven by variation in unobserved amenities (Sorkin, 2018), making firm fixed effects a good proxy for amenities. I then drop the top two and bottom two ventiles, since the variance of the firm fixed effects is much higher as we move into the tails of the distribution. Unobservable amenities can be reasonably assumed to be roughly constant within the remaining ventiles. I then re-estimate the firm specification defining the cell as a ventile of the distribution of wage fixed effects. Table A.2 reports the share of negative and significant discontinuities in each specification and the median lower and upper bound from this specification in Panel A, while empirical coverage rates are shown in Figure A.2. Estimates are again less precise than in the main specification, as there are fewer observations per cell, however there is still no evidence of tipping points in the composition of firms. The empirical coverage rate of the discontinuity is maximised at or close to zero in all years.

As an alternative proxy for firm-level amenities, I also use three-digit industries (Sorkin, 2018). The results from these specifications are in Panel B of Table A.2. The number of observations per cell in this case drops precipitously, which affects precision, but not the basic pattern of estimates. The distribution of confidence intervals for the estimated discontinuities is still centred around zero and the share of negative and significant discontinuities is never greater than 21 per cent.

A related concern is that the location of the tipping point might be purely due to technological factors which vary not by location, but by narrowly defined industry, or might on the contrary be driven primarily by historical or cultural factors that vary by

[20]The wage effects are estimated on the full sample of workers and firms subject to social security and included directly as a variable in the BHP from 1985, see Bellmann et al. (2020) for details of the estimation.

more narrowly defined location and not by industry. In any one of these cases, the true tipping dynamics will be attenuated by grouping together observations that in fact have different tipping points.

To address this concern, I repeat both the firm and industry estimations, this time defining the cell as either a single-letter industry (e.g. manufacturing, construction, or hotels and restaurants), or a local labour market (analogous to a Commuter Zone in the USA).[21] I again reports summary statistics on the share of negative discontinuities and the distribution of confidence intervals in Table A.2. When assuming that the location of the tipping point is common to firms or local industries in same single-letter industry (sector), shown in Panels C and D, the share of negative and significant cells is not much higher; the highest value attained is 0.4 in the case of firms in 1990. However, the confidence intervals in this case are relatively large and the empirical coverage rates for sector cells are still maximised near zero. When assuming that the location of the tipping point varies by local labour market, as shown in Panels E and F of Table A.2 the results are again in line with the main results. The share of negative and significant cells is not greater than 0.25 in any year. The confidence intervals are systematically centred on zero, even if they are somewhat wider, since there are more labour market cells than industry or location-by-industry type cells, reducing the number of observations in each labour market cell.

Of course, grouping firms within more narrowly defined cells does not formally rule out the possibility that there are firm-specific tipping points deriving from firm-specific amenities, and an alternative interpretation of the finding of no tipping points in the composition of firms will always be that preference spillovers simply lead to firm-specific tipping points which can't be detected using a threshold regression. Nevertheless, the evidence presented in this section on the absence of tipping points considering alternative proxies for firm-specific amenities allows us to conclude with some confidence that there are unlikely to be strong preference spillovers in the composition of workplaces by nativity. This is particularly so when one considers that when testing for tipping points by gender in the labour market, Pan (2015) found strong evidence of tipping dynamics when assuming a common tipping point to occupations grouped within quite coarse regional-skill cells. This is in spite of a long-standing body of research documenting differences in amenities across occupations going back to Lucas (1977) and Brown (1980) and clear gender differences in the preference for different amenities (Bell, 2020; Goldin, 2014a; Mas and Pallais, 2017). If preference spillovers were as important for the composition of workplaces by nativity as they appear to be for the gender composition of occupations, one could reasonably expect the threshold regression used here to detect tipping points in the composition of firms or

[21] I do not consider smaller aggregates in order to keep the sample size sufficiently large to estimate a threshold regression.

industries.

6 Conclusion

Tipping-like dynamics have been identified in neighbourhood composition, school enrolments or occupational composition, and have been argued to explain segregation in these different settings. This paper considered whether such tipping points could also contribute to explaining documented patterns of segregation between immigrants and natives across workplaces.

Under the assumption that the region-industry cells I use to group firms and industries adequately capture differences in amenities that determine the location of the tipping point, I have found relatively clear evidence against the existence of tipping points in the composition of both firms and local industries in Germany. The estimated discontinuities in net native workforce growth are typically small. Confidence intervals constructed using recently proposed inference methods for this type of problem typically allow me to rule out negative discontinuities greater than around 15 percentage points in magnitude for the case of firms and 25 percentage points in the case of industries.

Given the limited evidence of tipping points presented here, one can conclude that preference interactions are unlikely to be a leading explanation of observed workplace segregation, since tipping points are a necessary condition for the existence of preference spillovers. This is in contrast to residential markets and occupational composition, where preference spillovers are thought to be a leading explanation for observed patterns of segregation. Future research could therefore productively investigate what role alternative theoretical mechanisms, particularly the role of social networks in the job search process, play in explaining observed patterns of segregation.

References

Aldén, L., M. Hammarstedt, and E. Neuman (2015): "Ethnic segregation, tipping behavior, and native residential mobility," *International Migration Review*, 49, 36–69.

Andersson, F., M. García-Pérez, J. Haltiwanger, K. McCue, and S. Sanders (2014): "Workplace Concentration of Immigrants," *Demography*, 51, 2281–2306.

Andrews, I., T. Kitagawa, and A. McCloskey (2019): "Inference on Winners," Working Paper 25456, National Bureau of Economic Research, Boston.

——— (2021): "Inference after estimation of breaks," *Journal of Econometrics*, 224, 39–59.

Åslund, O. and O. N. Skans (2010): "Will I see you at work: Ethnic Workplace Segregation in Sweden, 1985-2002," *ILR Review*, 63, 471–493.

Aydemir, A. and M. Skuterud (2008): "The immigrant wage differential within and across establishments," *Industrial and Labor Relations Review*, 61, 334–352.

Banzhaf, H. S. and R. P. Walsh (2013): "Segregation and Tiebout sorting: The link between place-based investments and neighborhood tipping," *Journal of Urban Economics*, 74, 83–98.

Barth, E., B. Bratsberg, and O. Raaum (2012): "Immigrant wage profiles within and between establishments," *Labour Economics*, 19, 541–556.

Becker, G. S. and K. M. Murphy (2000): *Social Economics: Market Behaviour in a Social Environment*, Harvard University Press.

Bell, A. (2020): "Job Amenities and Earnings Inequality," Working paper.

Bellmann, L., B. Lochner, S. Seth, and S. Wolter (2020): "AKM effects for German labour market data," FDZ Methodenreport 01|2020, Institute for Employment Research (IAB), Nürnberg.

Böhlmark, A. and A. Willén (2020): "Tipping and the effects of segregation," *American Economic Journal: Applied Economics*, 12, 318–347.

Brown, C. (1980): "Equalizing differences in the labor market," *Quarterly Journal of Economics*, 94, 113–134.

Caetano, G. and V. Maheshri (2017): "School segregation and the identification of tipping behavior," *Journal of Public Economics*, 148, 115–135.

Card, D., A. Mas, and J. Rothstein (2008): “Tipping and the dynamics of segregation,” *Quarterly Journal of Economics*, 123, 177–218.

——— (2011): “Are mixed neighborhoods always unstable? two-sided and one-sided tipping,” in *Neighborhood and Life Chances: How Place Matters in Modern America*, ed. by H. B. Newburger, E. L. Birch, and S. M. Wachter, University of Pennsylvania Press, chap. 14.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016): “Referral-based job search networks,” *Review of Economic Studies*, 83, 514–546.

Elliott, G. and U. K. Müller (2007): “Confidence sets for the date of a single break in linear time series regressions,” *Journal of Econometrics*, 141, 1196–1218.

European Commission (2018): “Results of special Eurobarometer on integration of immigrants in the European Union,” Special Eurobarometer 469.

Ganzer, A., L. Schmidtlein, J. Stegmaier, and S. Wolter (2021): “The Establishment History Panel 1975–2019 (v2),” FDZ Datenreport 16|2020, Institute for Employment Research (IAB), Nürnberg.

Glitz, A. (2014): “Ethnic segregation in Germany,” *Labour Economics*, 29, 28–40.

Glover, D., A. Pallais, and W. Pariente (2017): “Discrimination as a self-fulfilling prophecy: evidence from french grocery stores,” *Quarterly Journal of Economics*, 132, 1219–1260.

Goldin, C. (2014a): “A grand gender convergence: Its last chapter,” *American Economic Review*, 104, 1091–1119.

——— (2014b): “A pollution theory of discrimination: Male and female differences in occupations and earnings,” in *Human Capital in History: The American Record*, ed. by L. P. Boustan, C. Frydman, and R. A. Margo, University of Chicago Press, 313–348.

Hansen, B. E. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” *Econometrica*, 64, 413–430.

——— (2011): “Threshold autoregression in economics,” *Statistics and its Interface*, 4, 123–128.

——— (2021): *Econometrics*, Princeton University Press.

Hellerstein, J. K. and D. Neumark (2008): “Workplace Segregation in the United States: Race, ethnicity, and skill,” *The Review of Economics and Statistics*, 90, 459–477.

Hjort, J. (2014): "Ethnic divisions and production in firms," *Quarterly Journal of Economics*, 129, 1899–1946.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019): "The allocation of talent and U.S. economic growth," *Econometrica*, 87, 1439–1474.

Imbens, G. W. and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.

Lee, D. S. and T. Lemieux (2010): "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48, 281–355.

Lubotsky, D. H. (2007): "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," *Journal of Political Economy*, 115, 820–867.

Lucas, R. E. B. (1977): "Hedonic wage equations and psychic wages in the returns to schooling," *American Economic Review*, 67, 549–558.

Mas, A. and A. Pallais (2017): "Valuing alternative work arrangements," *American Economic Review*, 107, 3722–3759.

Noiriel, G. (1988): *Le creuset français: Histoire de l'immigration (XIXe-XXe siècle)*, Paris: Éditions du Seuil.

OECD (2020): "International migration database," Data retrieved from International Migration Statistics, https://doi.org/10.1787/mig-data-en.

Pan, J. (2015): "Gender segregation in occupations: The role of tipping and social interactions," *Journal of Labor Economics*, 33, 365–408.

Sarvimäki, M. (2011): "Assimilation to a welfare state: Labor market performance and use of social benefits by immigrants to Finland," *Scandinavian Journal of Economics*, 113, 665–688.

Schelling, T. C. (1971): "Dynamic models of segregation," *Journal of Mathematical Sociology*, 1, 143–186.

——— (1978): *Micromotives and Macrobehavior*, New York: Norton.

Sorkin, I. (2018): "Ranking firms using revealed preference," *Quarterly Journal of Economics*, 133, 1331–1393.

Zheng, L. (2014): "Essays in Labor Economics," Ph.D. thesis, Cornell University.

A Supplementary tables and figures

Table A.1: Estimated discontinuities in firm net native workforce growth, medium firms

	(1) 1975	(2) 1980	(3) 1985	(4) 1990	(5) 1995	(6) 2000	(7) 2005
Tipping point	39.5	52.6	38.8	36.5	45.6	54.5	42.8
	(24.3)	(33.2)	(31.0)	(30.3)	(27.4)	(29.0)	(34.0)
Discontinuity ($\hat{\delta}$)	-0.9	-29.6	0.3	-19.3	-4.4	-4.2	7.1
	(46.4)	(51.0)	(40.4)	(48.4)	(60.0)	(48.8)	(60.9)
Median-unbiased $\hat{\delta}$	-5.1	-19.6	-11.2	-17.0	-5.2	10.3	14.6
	(40.2)	(51.4)	(47.0)	(46.2)	(51.3)	(38.5)	(54.2)
Share p-val. < 0.05	0.43	0.14	0	0.43	0.21	0.29	0.43
p-val. < 0.05 and $\hat{\delta} < 0$	0.21	0.07	0	0.21	0.14	0.14	0.21
Median LB, 95% CI	-24.4	-44.1	-26.1	-20.8	-30.9	-39.7	-9.3
Median UB, 95% CI	38.7	39.3	20.5	7.6	60.0	31.0	19.2
Cells	14	14	14	14	14	14	14
Avg. obs.	5331	5960	5929	6479	6644	8854	9286

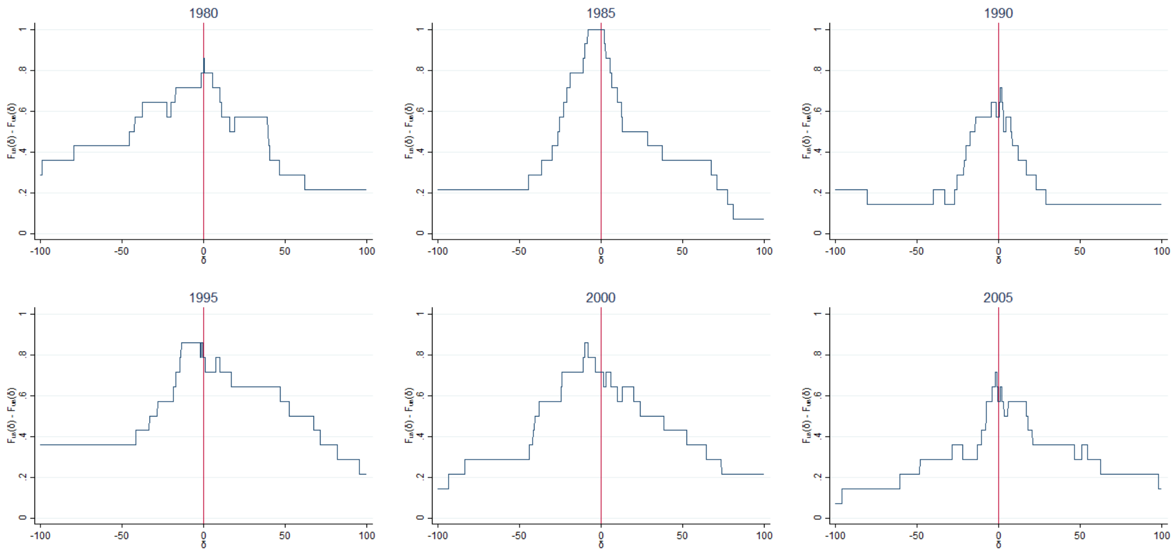
Note: The first section reports averages and standard deviations (in parentheses) of the estimated discontinuity $\hat{\delta}$ across region by industry-type cells. The second section reports measures of average significance, the third section reports the sample sizes. See text for details on estimation and inference procedures. The sample has been restricted to firms employing 10-49 workers in the base year.

Table A.2: Alternative cell definitions:

	(1) 1975	(2) 1980	(3) 1985	(4) 1990	(5) 1995	(6) 2000	(7) 2005
A: Wage FE cells (firms)							
p-val. < 0.05 and $\hat{\delta} < 0$			0.13	0.38	0.25	0.19	0.06
Median LB, 95% CI			-8.35	-11.55	-61.72	-21.30	-24.09
Median UB, 95% CI			25.68	30.78	30.79	15.08	20.99
wage FE cells			16	16	16	16	16
Avg. obs.			5796	6339	6170	7997	8279
B: Industry cells (firms)							
p-val. < 0.05 and $\hat{\delta} < 0$	0.17	0.21	0.12	0.15	0.16	0.18	0.11
Median LB, 95% CI	-51.74	-46.02	-46.01	-49.22	-44.98	-46.29	-28.44
Median UB, 95% CI	46.01	40.56	44.49	41.21	55.08	47.69	58.86
Industry cells	136	141	145	149	151	157	151
Avg. obs.	702	747	719	768	774	980	1064
C: Sector cells (firms)							
p-val. < 0.05 and $\hat{\delta} < 0$	0.13	0.27	0.13	0.40	0.07	0	0.07
Median LB, 95% CI	-21.46	-40.13	-28.88	-24.65	-26.13	-4.57	-17.25
Median UB, 95% CI	47.44	15.97	32.93	6.46	16.88	33.94	34.17
Sector cells	15	15	15	15	14	15	15
Avg. obs.	6409	7071	6997	7680	8390	10300	10766
D: Sector cells (industries)							
p-val. < 0.05 and $\hat{\delta} < 0$	0.15	0	0.08	0	0.08	0.15	0.17
Median LB, 95% CI	-3.96	-11.61	-29.30	-14.35	-41.34	-9.80	-21.97
Median UB, 95% CI	25.97	25.12	28.70	26.61	43.25	36.61	13.19
Sector cells	13	12	12	13	13	13	12
Avg. obs.	422	469	486	498	502	586	635
E: Labour market cells (firms)							
p-val. < 0.05 and $\hat{\delta} < 0$	0.08	0.15	0.13	0.18	0.21	0.13	0.18
Median LB, 95% CI	-35.98	-39.30	-28.81	-29.31	-28.14	-34.90	-29.79
Median UB, 95% CI	50.07	35.23	27.71	29.94	44.92	30.13	22.67
Labour market cells	39	39	39	39	39	39	39
Avg. obs.	2465	2720	2691	2954	3012	3962	4141
F: Labour market cells (industries)							
p-val. < 0.05 and $\hat{\delta} < 0$	0.13	0.25	0.13	0.16	0.08	0.22	0.03
Median LB, 95% CI	-34.02	-50.36	-38.82	-49.43	-30.47	-36.21	-19.78
Median UB, 95% CI	45.68	31.75	45.36	34.16	116.15	50.67	44.85
Labour market cells	32	32	30	32	37	37	35
Avg. obs.	168	178	190	199	180	210	218

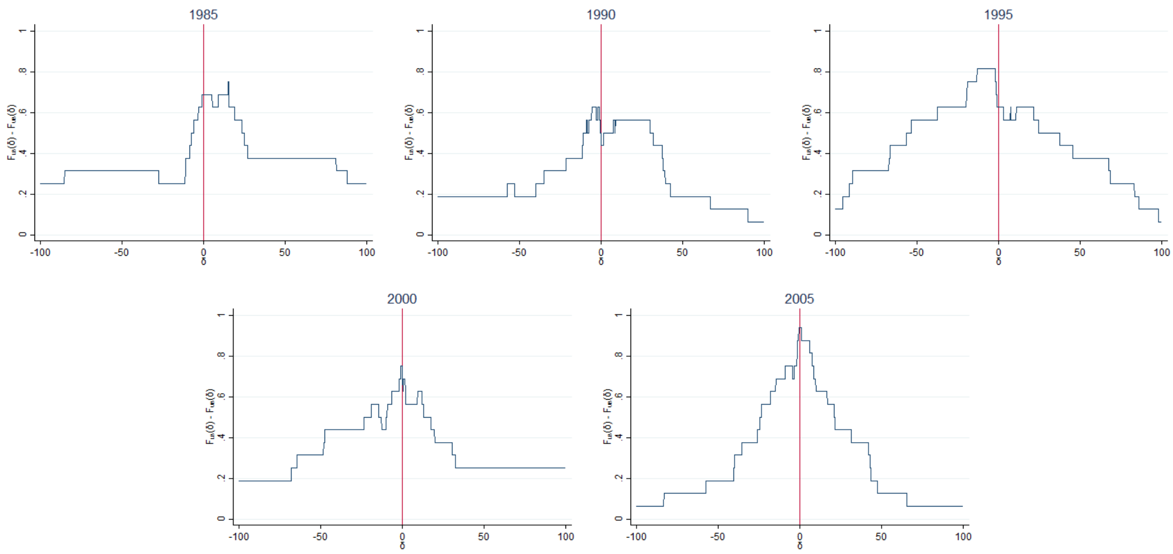
Note: Share of cells where the estimated discontinuity is negative and significant, and median bounds of 95 per cent confidence intervals. In panels A and B cells are defined as one-letter industries, in panels C and D, cells are defined as local labour markets. In panels A and C the unit of observation is the firm, in panels B and D the unit of observation is the local industry.

Figure A.1: Empirical coverage rates, medium-sized firms



Notes: The figures plot the empirical coverage rates for different values of $\delta \in [-100, 100]$. The empirical coverage rate is defined as the empirical CDF of the lower bound of the 95 per cent confidence interval for $\hat{\delta}$ minus the empirical CDF of the upper bound of the 95 per cent confidence interval for $\hat{\delta}$.

Figure A.2: Empirical coverage rates, firm fixed effect ventiles



Notes: The figures plot the empirical coverage rates for different values of $\delta \in [-100, 100]$. The empirical coverage rate is defined as the empirical CDF of the lower bound of the 95 per cent confidence interval for $\hat{\delta}$ minus the empirical CDF of the upper bound of the 95 per cent confidence interval for $\hat{\delta}$.

B Implementation details of inference procedure

This appendix sets out the detail of the threshold model that I estimate and defines the quantities necessary for the implementation of the inference procedures used, which are those developed in Andrews et al. (2019, 2021). The general model I estimate can be written as

$$Y_i = C_i'\beta + D_i'\delta\mathbf{1}(Q_i > \theta_0) + u_i, \tag{B.1}$$

where $C_i \in \mathbb{R}^d$ and $D_i \in \mathbb{R}^l$, with $1 \leq l \leq d$. This is very similar to the set-up considered by Andrews et al. (2021), only I allow for the possibility that the effect of only a sub-vector, D_i, of the full vector of control variables, C_i, varies when the variable Q_i crosses the threshold θ_0. While the results developed by Andrews et al. (2019, 2021) extend straightforwardly to this case, the definitions of various relevant quantities are slightly modified. Here I define the elements necessary to construct the estimators and confidence intervals defined by Andrews et al. (2019, 2021) when estimating the model defined in Equation (B.1).

Consider a finite parameter space Θ. Throughout I will define $\hat{\theta}_n$ as the NLS estimate of θ_0. For all $\theta \in \Theta$ define

$$X_n(\theta) = \begin{pmatrix} \left(\sum_{i=1}^n D_i D_i'\mathbf{1}\{Q_i \leq \theta\}\right)^{-1/2} \left(\sum_{i=1}^n D_i\eta_i\mathbf{1}\{Q_i \leq \theta\}\right) \\ \left(\sum_{i=1}^n D_i D_i'\mathbf{1}\{Q_i > \theta\}\right)^{-1/2} \left(\sum_{i=1}^n D_i\eta_i\mathbf{1}\{Q_i > \theta\}\right) \end{pmatrix} \tag{B.2}$$

where $\eta_i = D_i'\delta\mathbf{1}(Q_i > \theta_0) + u_i$. I assume that the threshold effect, δ, is small relative to sampling variability, which Elliott and Müller (2007) propose to model by assuming that $\delta = n^{-1/2}d$ for some $d \in \mathbb{R}$. Under this assumption, the arguments used in the proof of Proposition (1) in Elliott and Müller (2007) can be applied to show that $\hat{\theta}_n = \text{argmax}_{\theta\in\Theta}\|X_n(\theta)\| + o_p(1)$. This alternative (asymptotic) characterisation of $\hat{\theta}$ is useful to derive asymptotic confidence intervals for $\hat{\theta}_n$ or $\hat{\delta}(\hat{\theta}_n)$. Note furthermore that under the small threshold assumption and standard regularity conditions on the variable moments and covariances, it is straightforward to show that

$$\begin{aligned} X_n(\theta) \xrightarrow{d} X(\theta) = &\begin{pmatrix} \Sigma_{DD}(\theta)^{-1/2}\Sigma_{DDd}(\theta) \\ (\Sigma_{DD}(\bar{\theta}) - \Sigma_{DD}(\theta))^{-1/2})(\Sigma_{DDd}(\bar{\theta}) - \Sigma_{DDd}(\theta)) \end{pmatrix} \\ &+ \begin{pmatrix} \Sigma_{DD}(\theta)^{-1/2}G_D(\theta) \\ (\Sigma_{DD}(\bar{\theta}) - \Sigma_{DD}(\theta))^{-1/2})(G_D(\bar{\theta}) - G_D(\theta)) \end{pmatrix} \end{aligned}$$

where $\bar{\theta} = \sup(\Theta)$ and

$$
\begin{aligned}
n^{-1}\Sigma_{i=1}^{n} D_i D_i' \mathbf{1}\{Q_i \leq \theta\} &\xrightarrow{p} \Sigma_{DD}(\theta) \\
n^{-1}\Sigma_{i=1}^{n} D_i D_i' d\mathbf{1}\{Q_i > \theta_0\}\mathbf{1}\{Q_i \leq \theta\} &\xrightarrow{p} \Sigma_{DDd}(\theta) \\
n^{-1/2}\Sigma_{i=1}^{n} D_i u_i \mathbf{1}\{Q_i \leq \theta\} &\xrightarrow{d} G_D(\theta) \sim \mathcal{N}(0, \Sigma_{GD})
\end{aligned}
$$

Furthermore, define $Y_n(\theta) = e_j \sqrt{n}\hat{\delta}(\theta)$, where $\hat{\delta}(\theta)$ is the OLS estimate of δ after setting $\theta_0 = \theta$ and $e_j \in \mathbb{R}^l$ is the jth basis vector. Then, under the same standard regularity conditions as before, standard regression algebra can be used to show that

$$
Y_n(\theta) \xrightarrow{d} \mathcal{A}(\theta)^{-1}(\mathcal{B}(\theta) + \mathcal{C}(\theta)) \tag{B.3}
$$

where, extending the previous notation,

$$
\begin{aligned}
\mathcal{A}(\theta) &= \Sigma_{DD}(\bar{\theta}) - \Sigma_{DD}(\theta) - (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))\Sigma_{CC}(\bar{\theta})^{-1}(\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))' \\
\mathcal{B}(\theta) &= \Sigma_{DDd}(\bar{\theta}) - \Sigma_{DDd}(\theta) - (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))\Sigma_{CC}(\bar{\theta})^{-1}\Sigma_{CDd}(\bar{\theta})' \\
\mathcal{C}(\theta) &= G_D(\bar{\theta}) - G_D(\theta) - (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))\Sigma_{CC}(\bar{\theta})^{-1}G_C(\bar{\theta}).
\end{aligned}
$$

$X_n(\theta)$ and $Y_n(\theta)$ are therefore asymptotically normal. The asymptotic covariance matrices, $\Sigma_{XY}(\theta, \tilde{\theta})$ and $\Sigma_Y(\theta, \tilde{\theta})$ can be shown to be as follows:

$$
\Sigma_{XY}(\theta, \tilde{\theta}) = \begin{pmatrix} \Sigma_{DD}(\theta)^{-1/2}\mathrm{E}[G_D(\theta)\mathcal{C}(\tilde{\theta})']\mathcal{A}(\tilde{\theta})^{-1}e_j \\ (\Sigma_{DD}(\bar{\theta}) - \Sigma_{DD}(\theta))^{-1/2}(\mathrm{E}[G_D(\bar{\theta})\mathcal{C}(\tilde{\theta})'] - \mathrm{E}[G_D(\theta)\mathcal{C}(\tilde{\theta})'])\mathcal{A}(\tilde{\theta})^{-1}e_j \end{pmatrix} \tag{B.4}
$$

$$
\Sigma_{YY}(\theta, \tilde{\theta}) = e_j'\mathcal{A}(\theta)^{-1}\mathrm{E}[\mathcal{C}(\theta)\mathcal{C}(\tilde{\theta})']\mathcal{A}(\tilde{\theta})^{-1}e_j \tag{B.5}
$$

where

$$
\begin{aligned}
\mathrm{E}[G_D(\theta)\mathcal{C}(\tilde{\theta})'] =& \mathrm{E}[G_D(\theta)G_D(\hat{\theta})'] - \mathrm{E}[G_D(\theta)G_D(\tilde{\theta})'] \\
& - \mathrm{E}[G_D(\theta)G_C(\hat{\theta})']\Sigma_{CC}^{-1}(\Sigma_{DC}(\hat{\theta}) - \Sigma_{DC}(\tilde{\theta}))' \\
\mathrm{E}[\mathcal{C}(\theta)\mathcal{C}(\tilde{\theta})'] =& \mathrm{E}[G_D(\bar{\theta})G_D(\bar{\theta})'] - \mathrm{E}[G_D(\theta)G_D(\bar{\theta})'] \\
& + (\mathrm{E}[G_D(\theta)G_C(\bar{\theta})'] - \mathrm{E}[G_D(\bar{\theta})G_C(\bar{\theta})'])\Sigma_{CC}(\bar{\theta})^{-1}(\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\tilde{\theta}))' \\
& + (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))\Sigma_{CC}(\bar{\theta})^{-1}(\mathrm{E}[G_C(\bar{\theta})G_D(\tilde{\theta})'] - \mathrm{E}[G_C(\bar{\theta})G_D(\bar{\theta})']) \\
& + (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\theta))\Sigma_{CC}(\bar{\theta})^{-1}\mathrm{E}[G_C(\bar{\theta})G_C(\bar{\theta})']\Sigma_{CC}(\bar{\theta})^{-1} \\
& \quad \times (\Sigma_{DC}(\bar{\theta}) - \Sigma_{DC}(\tilde{\theta}))'.
\end{aligned}
$$

The conditional, unconditional, and hybrid confidence intervals and median-unbiased estimators defined in Andrews et al. (2019, 2021) can now be calculated for the model defined in Equation (B.1) by using the definitions of $X(\theta)$, $Y(\theta)$, $\Sigma_{YY}(\theta, \tilde{\theta})$, and $\Sigma_{XY}(\theta, \tilde{\theta})$ derived in this appendix in the definitions of the estimators and confidence intervals proposed by Andrews et al..

When implementing the estimators and confidence intervals defined by Andrews et al. (2021), we replace $X(\theta)$ with $\hat{X}_n(\theta)$, defined in Equation (B.2), where we substitute $\hat{\eta}_i = D_i'\hat{\delta}\mathbf{1}(Q_i > \hat{\theta}_n) + \hat{u}_i$ for η_i, letting, $\hat{\cdot}$ denote the NLS sample estimate of the parameters and errors defined in Equation (B.1). An estimate of $Y(\theta)$ is formed by taking the sample analogue of the limiting random variable in Equation (B.3), i.e. replacing the asymptotic matrices in the definitions of $\mathcal{A}(\theta)$, $\mathcal{B}(\theta)$, and $\mathcal{C}(\theta)$ by their sample analogues. Finally, to estimate the covariance matrices defined in Equations (B.4) and (B.5), I estimate $\mathrm{E}[G_D(\theta)G_D(\tilde{\theta})']$ using the heteroskedasticity-robust sample covariance matrix $n^{-1}\sum_{i=1}^{n} D_i D_i'\hat{u}_i^2\mathbf{1}\{Q_i \leq \min(\theta, \tilde{\theta})\}$, where $\hat{u}_i$ are again the NLS estimates of the errors defined in Equation (B.1).