Structural Estimation of Labor Adjustment Costs with Temporally Disaggregated Data and Gross Employment Flows

Fırat Yaman*

January 29, 2018

Abstract

Estimating labor adjustment costs is plagued by a variety of errors, many arising from data limitations. Most researchers have assumed that adjustment decisions are made at the firm level, that adjustment happens at the frequency at which a firm is observed (typically annually or quarterly), and that adjustment costs are incurred on net changes in employment. In this paper, I estimate a dynamic optimization model of labor adjustment of establishments based on data that permit 1) specifying any desired adjustment frequency, 2) using the micro unit of an establishment, 3) estimating the model based on net and on gross employment flows and 4) allowing for simultaneous hirings and separations. Results for adjustment costs depend crucially on the model specification. Only a monthly adjustment model yields positive cost parameters, while estimates from quarterly and annual adjustment models imply negative adjustment costs (that is, adjustment implies a gain rather than a loss). Estimating the model on net employment changes implies hiring and separation costs of more than four annual median salaries, while the model on gross changes implies costs on the order of two annual median salaries. Adjustment costs differ significantly between small and large establishments. However, a static specification of the model performs equally well as the dynamic model with respect to out-of-sample predictions.

JEL classification: C25; D22; J23

Keywords: Dynamic discrete choice; Adjustment costs; Labor demand; Temporal aggregation

*Department of Economics, City University London. Email: Firat.Yaman.1@city.ac.uk.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest: none.

1 Introduction

Estimating the magnitude and structure of labor adjustment costs has a long tradition in economics and has faced a number of challenges.[1] The first is the choice of the unit that makes hiring and separation decisions ("spatial" aggregation). Hamermesh (1989) showed that at the plant level long spells of inactivity are followed by lumpy labor adjustment – indicating either the presence of fixed costs of adjustment or non-differentiability at a net adjustment of zero; see, for example, Abel and Eberly (1994). Aggregated to the firm level, labor adjustment appears to be much more frequent and smooth, suggesting a convex adjustment cost function. Researchers who have been interested in the adjustment costs per se, rather than using them as a device for a better model fit, have employed plant or firm-level data over higher aggregates ever since. Second, a similar problem occurs with "temporal" aggregation. A plant or firm is more likely to change its employment in the long run. Thus, an infrequently observed plant is likely to exhibit no or few inactive periods with the same implications for cost estimates as in the spatial aggregation case, as demonstrated by Varejão and Portugal (2007) who use quarterly data. Third is the correct specification of the cost function. Hamermesh and Pfann (1996) rejected the hypothesis of convex adjustment costs in favor of other types of costs such as fixed costs, linear costs, and production disruption costs.[2] Forth is whether to use net or gross employment flows. When only the stock of employment at fixed intervals is reported, then net flows must be used. If adjustment costs are costs of hiring and separations, then gross flows should be used as in Abowd and Kramarz (2003) and Kramarz and Michaud (2010). Abowd et al. (1999) and Burgess et al. (2001) showed that net and gross flows differ considerably within an establishment, even within a skill group. Fifth is labor heterogeneity which is closely related to the previous point. All structural estimations of adjustment costs have specified a production function with one type of labor without acknowledging that in such a framework it makes no sense for an establishment to hire and fire within a period. Thus, adjustment costs in such a model can only be incurred over net changes. Goolsbee and Gross (2000) illustrated how this heterogeneity aggregation impacts on estimates of capital adjustment costs. Finally, a choice must be made concerning whether adjustment costs should be estimated using reported costs by firms and establishments or using a structural model. If the objective is to quantify adjustment costs, then the former is preferable and

[1] I examine labor adjustment, but the problems associated with estimating capital (or any factor) adjustment costs are almost identical. Hamermesh and Pfann (1996) have written an excellent survey for the literature until 1996 and addressed all of the following points.

[2] One or several of those components have been used by Abel and Eberly (1994); Abowd and Kramarz (2003); Aguirregabiria and Alonso-Borrego (2014); Asphjell et al. (2014); Bentolila and Bertola (1990); Bloom (2009); Cooper et al. (2015); Ejarque and Portugal (2007); Lapatinas (2009); Pfann and Palm (1993); Nilsen et al. (2007); Rota (2004); and Varejão and Portugal (2007).

was done by Abowd and Kramarz (2003) as well as Kramarz and Michaud (2010). Still, not all costs might be captured by establishment reports. In that case, or if interest lies in comparing models with and without adjustment costs, or in counter-factual policy evaluations, structural estimation is the method of choice.

In this paper, I employ the structural econometric approach to estimate labor adjustment costs. I parameterize labor adjustment costs in a model of dynamic programming where an establishment chooses its optimal employment given the state of the world and taking into account that its choice will affect its state vector in the next period. As such, my research is in the tradition of Aguirregabiria and Alonso-Borrego (2014); Cooper et al. (2015); Hamermesh (1992); and Rota (2004) to name but a few. Theoretically, my model accommodates the presence of hirings and separations within the same time period, without having to assume that costs are incurred only on net changes. To my knowledge, there is no other model of adjustment costs that does this. I show that estimates of adjustment costs depend crucially on this feature. Empirically, I demonstrate that assumptions about adjustment frequency and whether costs are incurred on net or on gross changes matter a lot for estimated adjustment costs. I do not claim that any one set of estimation results is correct. Rather, my results show that none of the estimated adjustment costs from structural labor demand models in the literature are credible. Thus, to estimate adjustment costs, the issue of adjustment frequency needs to be settled.

Based on the likelihood function derived from the dynamic programming model, I derive some hypotheses about how the estimated adjustment costs will change with respect to 1) different levels of time aggregation, 2) different levels of spatial aggregation, and 3) using net or gross flows of employees. I do not assume that labor is homogeneous. These hypotheses are tested empirically. Thus, my research addresses the points of temporal and spatial aggregation, net vs. gross flows, and labor heterogeneity, not offering a solution in terms of the "correct" level of aggregation or flows to be used, but highlighting how these choices matter intuitively and practically, and demonstrating this empirically.

I use linked employer-employee data from Germany which record annual sales figures for a sample of establishments, and, crucially, exact starting and ending days of employment spells for each employee (subject to paying social security contributions) who works in one of the establishments. This allows me to assume any employment adjustment frequency, from daily to annual, and to observe the gross flows of employees within any period.

I build a model of linear hiring and separation costs which accommodates churning, the observation of simultaneous hirings and separations, and estimate it for different aggregation levels, and separately for adjustment costs defined over gross versus net flows. I show that the most important factor contributing to higher adjustment cost estimates are the number of observations of inactivity. The economically most sensible estimates of adjustment costs are obtained for a model of 1) monthly adjustment on 2) gross employee flows and figure around 40,000 Euros per hire and per separation. Without an external source concerning the magnitude of adjustment costs there is nothing to compare these estimates against.[3]

Despite the above mentioned difficulties economists should be interested in estimating adjustment costs. Macroeconomists will be interested in the effect of adjustment costs on unemployment and the labor market effects of the business cycle; see, for example Bentolila and Bertola (1990) and Hopenhayn and Rogerson (1993). Policymakers will be interested in the role of adjustment costs when considering labor market interventions such as wage subsidies in a recession (such as the German *Kurzarbeit* program), and economists working in industrial organization should know the right specification of a firm's objective function for purposes such as estimating a production function. Labor economists are, I presume, intrinsically interested.

2 Model

Establishments are assumed infinitely-lived, maximizing the present value of current and all expected future profits, and being price takers in product and factor markets. Thus, the only choice variables are how many workers to hire, and from how many to separate. I assume that the establishment cannot observe the quality of a worker it hires, but it observes the quality of its existing workers. However, it cannot pay individual-specific wages. If the establishment decides to separate from workers, then it will do so from the ones with the smallest (expected) marginal productivities. Hiring or separating from workers incurs adjustment costs. The general formulation of an establishment's problem is given by

$$V(\mathbf{x}_t, \mathbf{a}_{t-1}) = \max_{h_t \geq 0, f_t \geq 0} \pi^e(\mathbf{x}_t, \mathbf{a}_{t-1}, h_t, f_t) - C(h_t, f_t) + \beta \mathbb{E}_{\mathbf{x}_{t+1}, \mathbf{a}_t | \mathbf{x}_t, \mathbf{a}_{t-1}, h_t, f_t} V(\mathbf{x}_{t+1}, \mathbf{a}_t)$$

[3]Unfortunately, but not unsurprisingly, estimates in the literature have a very wide range. For Germany, Mühlemann and Pfeifer (2013) reported hiring costs between 4,000 and 6,000 Euros per hire. However, Hamermesh and Pfann (1996) cited some accounting studies which find adjustment costs of one year of payroll costs for the average worker. More on this in the discussion section.

where $\mathbf{x}_t$ is a vector characterizing the state that is relevant for the establishment's decision, $\mathbf{a}_{t-1}$ is the vector (and set) of the productivities of its workers before any hiring and separation decisions are made, π^e is the expected profit function, and h_t and f_t are the (non-negative) number of hires and separations chosen by the establishment. The next period is discounted at rate β, and C is the adjustment cost function. Expectations are taken over the future state and over labor productivities conditional on the current state and labor productivity (before labor adjustment decisions are made). Here I define labor productivity as anything that increases establishment revenue per worker, including factors such as demand shocks and changes in the output price. The timing of the model is as follows: at the beginning of period t the establishment observes the productivities of its workers $\mathbf{a}_{t-1}$ and other relevant state variables $\mathbf{x}_t$. It then decides on h_t and f_t. After adjustment, the productivities of retained workers and new hires become known to the establishment. The current profits are expected, as the establishment only has an expectation over the quality of its new hires and the evolution of the quality of its existing workers before the hiring is made. The problem becomes dynamic via the link between the inherited workers, hires, separations and the new stock of workers. Let $\mathbf{h}_t$, $\mathbf{f}_t$, and $\boldsymbol{\eta}_t$ be the set of productivities of new hires, employees which the establishment separates from, and employees who leave the establishment without a separation decision of the establishment (such as voluntary quits or retirements) respectively. Let $\boldsymbol{\varepsilon}_t$ be the set of random shocks to the productivities of workers. The evolution of the stock and quality of the establishment's employment is given by

$$\hat{\mathbf{a}}_t = (\mathbf{a}_{t-1} \setminus \mathbf{f}_t \setminus \boldsymbol{\eta}_t) \cup \mathbf{h}_t \tag{1}$$

and

$$\mathbf{a}_t = \hat{\mathbf{a}}_t + \boldsymbol{\varepsilon}_t \tag{2}$$

In words, from the existing set of workers, those who the establishment separates from and those who quit are removed and new hires are added. Then each worker receives a productivity shock.

If the profit function and adjustment cost function were differentiable everywhere, then it would be easy to derive an Euler equation and to estimate the adjustment cost parameters with generalized method of moments. Evidence against an everywhere differentiable cost function is quite strong; see, for example, Hamermesh and Pfann (1996). Employment changes are relatively rare, and tend to be lumpy when they do occur, suggesting the presence of either fixed costs or non-differentiability at the point of no adjustment.

Moreover, at least for small establishments, assuming a continuously adjustable level of employment is also problematic, in particular if the empirical model is set up in a way that non-activity translates into higher adjustment cost estimates, thus confounding adjustment costs with indivisibilities of labor.[4] I model the adjustment cost function to be linear in hires and separations, thus introducing a non-differentiability at zero adjustment and consequently a range for the state space that will make inactivity the optimal choice of an establishment as in Abel and Eberly (1994). An Euler equation approach with non-differentiability is still possible as demonstrated in Aguirregabiria (1997) and Cooper et al. (2010), though the requirements for consistent estimation of this Euler equation are unlikely to be met. The main problem is the endogeneity of the moment of adjustment; see Aguirregabiria (1997) for a detailed exposition.

Let the profit function be given by

$$\pi(\mathbf{a}_t, w_t) = S(\mathbf{a}_t) - w_t \ell_t$$

where w_t is the average wage paid in the establishment, ℓ_t is the number of workers – the length of vector $\mathbf{a}_t$ – and S denotes sales net of the cost of intermediate inputs. To derive an estimable model, I have to impose some structure on the sales function. The structural estimation literature has assumed a specific production function at this point. I assume that the sales function can be factored into a part that depends on ℓ_t only and a part that depends on a statistic of $\mathbf{a}_t$:

$$S(\mathbf{a}_t) = A(\mathbf{a}_t) f_\ell(\ell_t)$$

As a result, I can write A, which is not observed, as a function of sales and number of workers, which are both observed. This assumption rules out some classes of production functions, but is still more general than specifying a fully parametric production function. From equations 1 and 2, $\mathbf{a}_t$ depends on $\mathbf{a}_{t-1}$, hires, separations, and unobserved shocks. Therefore, so does A:

$$A_t = f_A(A_{t-1}, h_t, f_t, \varepsilon_t)$$

where ε is an independent random variable. Thus, I can characterize the current sales as a function of current hiring and separation choices (and therefore current employment), previous period's employment,

[4]This point is also made in Lapatinas (2009) who models the choice of employees as a discrete choice.

previous period's sales, and an unobservable (for the researcher) component ε:

$$
\begin{aligned}
S_t =& A_t f_\ell(\ell_t) \\
=& f_A(A_{t-1}, h_t, f_t, \varepsilon_t) f_\ell(\ell_t) \\
=& f_A\left(\frac{S_{t-1}}{f_\ell(\ell_{t-1})}, h_t, f_t, \varepsilon_t\right) f_\ell(\ell_t) \\
=& f_S(S_{t-1}, \ell_{t-1}, h_t, f_t, \varepsilon_t, \ell_t)
\end{aligned}
$$

The state vector of the establishment when considering its employment decision thus consists of S_{t-1}, ℓ_{t-1} and the current wage paid to its employees w_t. Recall that the establishment makes decisions based on the expected sales, that is before it can observe ε_t. Nothing rules out that an establishment might hire and separate simultaneously. For example, observing low employee productivity, an establishment might choose to replace an unproductive with a productive worker, since by doing so it can increase sales (through A) even though it holds its number of employees constant. Denoting the observable state variables $\mathbf{x_t} = (w_t, S_{t-1}, \ell_{t-1})$, we can write the establishment's optimization problem as

$$
V(\mathbf{x_t}) = \max_{h_t \geq 0, f_t \geq 0} \pi^e(\mathbf{x_t}, h_t, f_t) - C(h_t, f_t) + \beta \mathbb{E}_{\mathbf{x_{t+1}} | \mathbf{x_t}, h_t, f_t} V(\mathbf{x_{t+1}}) \tag{3}
$$

This problem has a unique fixed point V^*; see Theorem 9.6 in Stokey and Lucas (1989). I denote the optimal decision rule to this problem by $\delta(\mathbf{x_t}) = [h^*(\mathbf{x_t}), f^*(\mathbf{x_t})]$. This decision rule is guaranteed to be a function (rather than a correspondence) if π is concave and C is of the form

$$
C(h_t, f_t) = \tau^{fix} \mathbb{1}(h_t > 0 \quad \vee \quad f_t > 0) + \tau^+ h_t + \tau^- f_t \qquad h_t \geq 0, \quad f_t \geq 0
$$

where τ^{fix} is a fixed cost of hiring and/or separations, $\mathbb{1}$ is the indicator function, and τ^+ and τ^- are "unit" costs of hiring and separations respectively. The function $\pi^e - C$ is K-concave and the resulting decision rule is of the (S,s) type; see Scarf (1959) and Aguirregabiria (1999) for this result. Following the steps in Aguirregabiria (1999), the decision space of the firm can be split up into four discrete areas: Let H be the discrete choice of $h > 0$ and $f = 0$, F the choice of $h = 0$ and $f > 0$, P (for put) the choice of $h = 0$ and $f = 0$, and B (for both) the choice of $h > 0$ and $f > 0$. Denote the choice set of exhaustive and mutually exclusive choices by $D = \{H, F, P, B\}$. Let d be the discrete choice of an establishment: $d \in D$. With each of these choices, there is an associated optimal level of hires and separations. For example, for $d = H$, we would constrain $f = 0$, but h would be the solution to the problem

$$V^H(\mathbf{x_t}) = \max_{h_t > 0} \pi^e(\mathbf{x_t}, h_t, f_t = 0) - C(h_t, f_t = 0) + \beta \mathbb{E}_{\mathbf{x_{t+1}}|\mathbf{x_t}, h_t, f_t = 0} V(\mathbf{x_{t+1}})$$

Let $\delta^d(\mathbf{x_t}) = \left[h^d(\mathbf{x_t}), f^d(\mathbf{x_t}\right])$ denote the solutions for h and f in the discrete regime d. For a current choice d, next period's value function itself can be split into the four discrete choices so that

$$\mathbb{E}_{\mathbf{x_{t+1}}|\mathbf{x_t}, d_t} V(\mathbf{x_{t+1}}) = \mathbb{E}\left[\max_{j \in D}\left(V^j(\mathbf{x_{t+1}})\right)|\mathbf{x_t}, d_t\right]$$

The "discrete" version of the firm's problem in equation 3 is thus given by

$$V(\mathbf{x_t}) = \max_{d_t \in D} \pi^e(\mathbf{x_t}, \delta^d(\mathbf{x_t})) - C(\delta^d(\mathbf{x_t})) + \beta \mathbb{E}\left[\max_{j \in D}\left(V^j(\mathbf{x_{t+1}})\right)|\mathbf{x_t}, d_t\right] \tag{4}$$

and the solution to this problem is denoted $d_t^*(\mathbf{x_t})$.

To summarize, the establishment considers a discrete action (and the associated optimal h and f). In doing so, it forms expectations of the future state given its current state and this discrete action. It knows that in the next period it will again choose among the best of the four discrete options (and the associated optimal h and f). Finally, to make this model estimable, I decompose expected sales into the empirical conditional expectation and its deviation from it. Following again Aguirregabiria (1999) I model

$$\pi^e\left[\mathbf{x_t}, \delta^d(\mathbf{x_t})\right] = \mathbb{E}\left[\pi^d(\mathbf{x_t})\right] + \varepsilon_t^{d,\pi}$$

where the first term is the expected profit under discrete choice d, and $\varepsilon_t^{d,\pi}$ captures unobservable shocks to the profits under this discrete choice. Note that the establishment observes both components, but the researcher only estimates the first with a sample average. By construction $\mathbb{E}\left(\varepsilon_t^{d,\pi}|\mathbf{x}_t\right) = 0$. Similarly, let

$$C\left[\delta^d(\mathbf{x_t})\right] = \mathbb{E}\left[C^d(\mathbf{x_t})\right] + \varepsilon_t^{d,C}$$

so the adjustment costs are based on the expected adjustments under the choice d and an unobservable component $\varepsilon_t^{d,C}$. To simplify the notation from here on I define

$$u(\mathbf{x_t}, d_t) \equiv \mathbb{E}\left[\pi^d(\mathbf{x}_t)\right] - \mathbb{E}\left[C^d(\mathbf{x_t})\right]$$

$$\varepsilon_t^d \equiv \varepsilon_t^{d,\pi} + \varepsilon_t^{d,C}$$

$$\boldsymbol{\varepsilon_t} = (\varepsilon_t^H, \varepsilon_t^F, \varepsilon_t^P, \varepsilon_t^B)$$

With $\mathbf{x_t}$ containing three components, and having another set of four variables in $\boldsymbol{\varepsilon_t}$, this is a formidable problem to solve. Rust (1987) showed that under some assumptions the dimensionality of this problem can be greatly reduced by eliminating the $\boldsymbol{\varepsilon_t}$ from the state space. The following standard assumptions need to be made:

1. The shocks ε_t^d are independent across alternatives, across time, and follow an extreme value type I distribution (we will see that we will not need to normalize the scale parameter).
2. The distribution of the future observable state variables is independent of $\boldsymbol{\varepsilon_t}$ conditional on the current state variables and the discrete choice, that is $F(\mathbf{x_{t+1}}|d_t, \mathbf{x_t}, \boldsymbol{\varepsilon_t}) = F(\mathbf{x_{t+1}}|d_t, \mathbf{x_t})$.

The second assumption is known as the conditional independence assumption. It allows to integrate out the shocks from equation 4. Denoting $\bar{V}(\mathbf{x_t}) = \int_\varepsilon V(\mathbf{x_t}, \boldsymbol{\varepsilon_t}) \mathrm{d}G_\varepsilon(\boldsymbol{\varepsilon_t})$, and taking expectations with respect to $\boldsymbol{\varepsilon}$ I can derive

$$\int_\varepsilon V(\mathbf{x_t}, \boldsymbol{\varepsilon_t}) \mathrm{d}G_\varepsilon(\boldsymbol{\varepsilon_t}) = \int_\varepsilon \max_{\mathrm{d_t}} \left(\mathrm{u}(\mathbf{x_t}, \mathrm{d_t}) + \varepsilon_\mathrm{t}^\mathrm{d} + \beta \mathbb{E}_{\mathrm{x_{t+1}}, \varepsilon_{\mathrm{t+1}}} \left(\mathrm{V}(\mathbf{x_{t+1}}, \boldsymbol{\varepsilon}_{t+1} | \mathbf{x_t}, \mathrm{d_t})\right) \mathrm{d}G_\varepsilon(\boldsymbol{\varepsilon_t}\right)$$

$$\bar{V}(\mathbf{x_t}) = \int_\varepsilon \max_{d_t} \left(u(\mathbf{x_t}, d_t) + \varepsilon_t^d + \beta \mathbb{E}_{x_{t+1}} \left(\bar{V}(\mathbf{x_{t+1}} | \mathbf{x_t}, d_t)\right)\right) \mathrm{d}G_\varepsilon(\boldsymbol{\varepsilon_t}) \tag{5}$$

With assumption 1) I get

$$\bar{V}(\mathbf{x_t}) = \log\left(\sum_d \exp\left(u(\mathbf{x_t}, d_t) + \beta \mathbb{E}_{\mathbf{x_{t+1}}} \left(\bar{V}(\mathbf{x_{t+1}} | \mathbf{x_t}, d_t)\right)\right)\right)$$

and choice probabilities

$$P(d_t | \mathbf{x_t}) = \frac{\exp\left(v(d_t, \mathbf{x_t})/\sigma\right)}{\sum_{j \in D} \exp\left(v(j, \mathbf{x_t})/\sigma\right)} \tag{6}$$

with

$$v(d_t, \mathbf{x_t}) \equiv u(d_t, \mathbf{x_t}) + \beta \mathbb{E}(\bar{V}(\mathbf{x_{t+1}} | \mathbf{x_t}, d_t))$$

being the choice-specific value function and σ being the scale parameter of the type I extreme value distribution. In the framework of dynamic discrete choice models, the scale parameter is not identified if all additive terms in u are multiplied by parameters to be estimated. As we will see, in this model u will contain a term (the wage bill) that allows identification of σ.

Rust (1987) showed that equation 5 is a contraction mapping and suggests a nested fixed point algorithm to estimate this model. Aguirregabiria and Mira (2010) have provided a discussion of solution algorithms. Still, the computational burden is quite great, since one has to solve for $\bar{V}$ at every step of the likelihood iteration. I follow the exposition in Aguirregabiria and Mira (2010) to simplify the computation of the likelihood function. Let the expected profit function be given by

$$\mathbb{E}\left[\pi^d(\mathbf{x_t})\right] = \mathbb{E}(S_t | \mathbf{x_t}, d_t) - w_t \mathbb{E}(\ell_t | \mathbf{x_t}, d_t)$$

Sales, for reasons that I will explain later, will be measured with error. Furthermore, consider an adjustment cost function of the form

$$C(h_t, f_t) = \tau^+ h_t + \tau^- f_t$$

with expected adjustment costs conditional on $\mathbf{x}$ and the discrete choice d

$$\mathbb{E}(C^d(\mathbf{x_t})) = \tau^+ \mathbb{E}(h_t | \mathbf{x_t}, d_t) + \tau^- \mathbb{E}(f_t | \mathbf{x_t}, d_t)$$

I can thus write

$$u(\mathbf{x_t}, d_t) = \mathbf{z}(\mathbf{x_t}, d_t)\theta_u'$$

with

$$\mathbf{z}(\mathbf{x_t}, d_t) = [\mathbb{E}(S_t | \mathbf{x_t}, d_t) \quad w_t \mathbb{E}(\ell_t | \mathbf{x_t}, d_t) \quad \mathbb{E}(h_t | \mathbf{x_t}, d_t \in \{H, B\}) \quad \mathbb{E}(f_t | \mathbf{x_t}, d_t \in \{F, B\})] \tag{7}$$

and

$$\theta_u = \begin{pmatrix} 1/\sigma_s & -1 & -\tau^+ & -\tau^- \end{pmatrix}$$

Here σ_s is included as distinct from σ due to the measurement error in sales.[5] In the appendix I show – using a result from Hotz and Miller (1993) on expressing $\mathbb{E}(\varepsilon^{d_t}|d_t,\mathbf{x_t})$ in terms of choice probabilities and closely following the steps in Aguirregabiria and Mira (2010) – that the choice specific value function can be written as

$$v(\mathbf{x_t},d_t) = \tilde{\mathbf{z}}(\mathbf{x_t},d_t)\theta_u' + \tilde{\mathbf{e}}(\mathbf{x_t},d_t)\theta_e'$$

where the expressions for $\tilde{\mathbf{z}}$, $\tilde{\mathbf{e}}$ and θ_e are in the appendix. Furthermore, I can derive the following recursive equation[6]

$$\mathbf{W}(\mathbf{x}) = \sum_{d\in D} P(d|\mathbf{x}) \times \left([\mathbf{z}(\mathbf{x},d),\mathbf{e}(\mathbf{x},d)] + \beta F(\mathbf{x}'|\mathbf{x},d)\mathbf{W}(\mathbf{x}')\right) \tag{8}$$

where $\mathbf{W}(\mathbf{x})$ is defined in the appendix. $\mathbf{W}(\mathbf{x})^*$ is the unique solution to equation 8 and is needed to compute $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{e}}$ and ultimately the choice probabilities in equation 6.

Note how this formulation greatly simplifies the estimation once one has an initial estimate of the choice probabilities P and the state transition probabilities F. Since the last equation does not contain any parameters to be estimated, the vector $\mathbf{W}$ needs to be solved only once rather than at each step of the maximization algorithm.

3 Estimation

The estimation proceeds in three steps. First, I need to obtain estimates of the state transition probabilities $F(\mathbf{x}'|\mathbf{x},d)$ and of the choice probabilities $P(d|\mathbf{x})$. Second, I solve for the unique fixed point of equation 8. Third, I maximize the likelihood function based on the choice probabilities in equation 6.

In the data, I observe sales, employment, wages paid to every worker, the number of hires and the number of separations. I assume every worker at any point of time is paid the same wage within an establishment, which I take to be the median wage in the establishment rather than the mean due to some right-censoring issues of the wage data. I discretize the state space into an array of $\omega_w \times \omega_S \times \omega_\ell$ points, where I choose $\omega_w = 14$ for the wage, $\omega_S = 20$ for sales, and $\omega_\ell = 31$ for employment, creating thus an array of 8,680 points. I choose the cell boundaries for w and S to have an equal number of observations in each cell. For ℓ, I choose a fine discretization for small ℓ, and wider intervals for higher levels of employment. For each of

[5] I defer the discussion about sales to the sales section.

[6] The steps of the derivation are in the appendix.

these points and each discrete choice, I calculate $P(d|\mathbf{x})$ by a nearest-neighbor estimator. In particular, I first normalize the state vector (w,S,ℓ) to a mean of zero and an identity variance matrix. Call the normalized data $(\bar{w},\bar{S},\bar{\ell})$ and the normalized state point $\bar{\mathbf{x}}$. For each normalized $\bar{\mathbf{x}}$, I calculate the Euclidean distance of this point from the normalized data observations, and choose the k observations with the smallest distance. The relative frequencies of the discrete choices among those k observations are used as estimates for P. I follow Pagan and Ullah (1999) in choosing $k=\sqrt{n}$. The transition probability array $F(\mathbf{x}'|\mathbf{x},d)$ is estimated in the same vein, but only among observations with the discrete choice of interest.

The components of the vector $\mathbf{z}(\mathbf{x},d)$ are also estimated by a nearest-neighbor algorithm. I estimate $\mathbb{E}(S|\mathbf{x},d)$, $\mathbb{E}(\ell|\mathbf{x},d)$, $\mathbb{E}(h|\mathbf{x},d)$, and $\mathbb{E}(f|\mathbf{x},d)$ as the average sales, average employment, average hires and average separations among the k observations with discrete choice d and the smallest distance to the normalized state point $\bar{\mathbf{x}}$.

Having all these objects in place, I start with an initial guess for the vector $\mathbf{W}(\mathbf{x}')$, to iterate on equation 8 until convergence is achieved. Then I calculate $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{e}}$ for each observation. The parameter vector for the likelihood in equation 6 can be written as $\theta_u/\sigma=\tilde{\theta}_u=(1/(\sigma_s\sigma) \quad -1/\sigma \quad -\tau^+/\sigma \quad -\tau^-/\sigma)$ and $\tilde{\theta}_e=(1 \quad \mu/\sigma)$. The parameter μ/σ cannot be identified, since the corresponding element in $\mathbf{e}(\mathbf{x},d)$ is equal to one for each choice, and the likelihood requires differences in variables across choices. The remaining parameters are estimated by maximizing the log likelihood function of the sample. The costs of hiring and separating from one worker are given by τ^+ and τ^-. The parameter σ is the scale parameter of the type I extreme value distribution, and determines the variance of the distribution as $\frac{\pi^2}{6}\sigma^2$. It gives a measure of how variant the distribution needs to be to account for the hiring and separation heterogeneity of otherwise similar establishments and/or to account for the choice of d of an establishment which based on its choice specific value functions $v(x,j)$ should have chosen a different alternative. The stronger the model is in predicting the observed choices of the sample establishments, the smaller this variance should be. It is thus maybe comparable in spirit to the mean squared error in a linear regression model.

The reader might have noticed a certain inconsistency in parametrically estimating choice probabilities (call them P_2) in equation 6 relying on the non-parametrically estimated initial choice probabilities (P_1). I follow the methodology in Aguirregabiria and Mira (2002) who show that P itself is a fixed point in the sense that iterating on P will lead to a unique solution of the recursive function of P. Thus, I use P_2 based on the estimated parameters θ, to feed them into a renewed computation of $\mathbf{W}$. Then this $\mathbf{W}$ is used to obtain

P_3, and the procedure can be repeated any number of times or until convergence in the parameter vector is achieved. In the estimations this was the case after four iterations.

Given that the final likelihood maximization uses many constructed variables based on a state-space discretization and non-parametric estimation, I do not calculate standard errors of the estimates. Depending on the exact specification, obtaining one set of results on a high-performance computer using 20 nodes takes a few hours. Bootstrapping the standard errors with 500 replications would take me a few weeks. Instead of calculating standard errors, I evaluate the performance of my model by out-of-sample predictions.

3.1 Identification

The likelihood function I am maximizing is the likelihood function of a multinomial logit model, and has a unique maximizer, so the model is identified in that sense. In this subsection I am interested in the intuition of what will determine sign and magnitude of the structural coefficients. I estimate the parameters of this model freely. The data can invalidate this model in a variety of ways. In particular, common sense would require $1/(\sigma_s\sigma) > 0$, $-1/\sigma < 0$, $-\tau^+ < 0$, and $-\tau^- < 0$. That is, sales increase the value of a firm, the scale parameter is positive, paying wages decreases the value of the firm, and labor adjustment is indeed costly.[7] What do the data need to be like to yield these signs? The intuition can be captured – and the exposition simplified – if in this section we ignore the continuation value part of the value function. The likelihood of choosing the observed choice c out of a set D in a multinomial logit model is

$$\begin{aligned} P(d^* = c) =& \frac{\exp(x_c\beta')}{\sum_{a\in D}\exp(x_a\beta')} \\ =& \frac{1}{1+\sum_{a\neq c}\exp((x_a - x_c)\beta')} \end{aligned}$$

This likelihood will be the greater the smaller each component of $(x_a - x_c)\beta'$ is. Thus, if the coefficient on sales is to be positive, the sales value among the chosen alternative needs to exceed the sales value among the non-chosen alternatives sufficiently frequently. In general, the chosen option should dominate the non-chosen alternatives with respect to sales. The reverse is required for the coefficient on the wage bill. Ceteris paribus the wage bill paid under the chosen option needs to be lower than the wage bill under the non-chosen options sufficiently often to guarantee a negative sign on the associated parameter. Consider the

[7] To borrow from Tolstoy, there is only one way in which the model can be "correct", but many ways for it to be "wrong".

hiring cost parameter, so that $x_d = \mathbb{E}(h_t|\mathbf{x}_t, d_t)$. Suppose the chosen action is $c = P$, that is to neither hire nor separate from any worker. x_c in this case will be zero (since no one was hired), but x_a will be either zero (if $a = F$) or a positive number (if $a = H$ or $a = B$). Thus, $(x_a - x_c) \geq 0$. The same can be established for the separation cost parameter. This will decrease the associated coefficient (τ^+) to increase the likelihood. Contrast this with the choice of both hiring and separating, that is $c = B$. For the hiring cost parameter, two of the alternatives (P and F) will result in $(x_a - x_c) \leq 0$ (since $x_a = 0$ and $x_c = x_B > 0$) while for the alternative H the sign of $(x_a - x_c)$ is not known a priori. Thus, unless x_H is always much greater than x_B, this choice will tend to increase the hiring cost parameter. A similar argument can be made for the separation cost parameter. Finally, the choice H will tend to increase the hiring cost and decrease the separation cost parameter, and the reverse will be true for the choice F. Thus the more inactivity we see in the sample, the smaller the parameters $-\tau^+/\sigma$ and $-\tau^-/\sigma$ will be (and the greater τ^+ and τ^- will be provided they are positive). Conversely, having many B choices in the sample will increase the same parameters. In order to obtain positive τ^+ and τ^- we should thus have a sufficient number of inactive observations in our sample. This highlights quite clearly the issues of temporal and of size aggregation. First consider time aggregation. In practice, the shorter the intervals at which firms are assumed to revise their employment, the more likely they are to be inactive. Any hiring or separation during a month will also be a hiring or separation during the year, but a hiring or separation during the year could still mean eleven months of inactivity compared to one month of activity. As for size aggregation, a firm with thousands of employees is much more likely to be active than a small establishment. Treating the large firm as one entity – with one decision maker for hires and separations – and observing this entity constantly in employment adjustments should lead us to infer that it hiring and separating cannot be too costly, at least for small adjustments.

4 Data

The data are linked employer-employee data from Germany (LIAB longitudinal model version 3), covering the years 1993 to 2007. The data have a survey based employer side, and an administrative employee side. The employer survey is conducted annually through in-place interviews. The sampling unit is an *establishment* (in German *Betrieb*), not a firm as a legal entity. Roughly an establishment is a spatial and commercial unit. A firm might thus have many establishments. The population are all establishments with at least one employee subject to social security contributions. The sampling method is stratified random sampling, with larger establishments (in terms of employment) being oversampled. The survey collects information on annual sales, expenditures on intermediate inputs, employment, investment and many other areas. The employee side of the data comes from the administrative records of the German Employment Agency. Every employee subject to social security contributions must be reported by the employing establishment to the Agency for the purpose of computing and collecting social security contributions. As such, the establishment must also report the exact salary paid to the employee. From this administrative data, I know the beginning and the end date of employment spells of any employee employed by any of the establishments surveyed by the establishment panel. More information on the data can be found in Jacobebbinghaus (2008).

The main advantage of this data is the accurate observance of all employment flows. This is what allows me to use gross rather than net flows and what allows me to include the discrete option of simultaneous hiring and separation (in addition to only hiring and only separation) to the choice set. I outline now how I clean the data:

1. The establishment survey reports the stock of employees paying social security for the 30th of June. I compare this reported number to the stock of social security paying employees on the same day from the administrative records. Ideally both numbers should be the same. I drop establishment-year observations where the discrepancy is too large.[8] I also drop establishment-years which report a share of social security paying employees among all employees greater than 100%. Finally I drop establishment-years for which this share is 50% or less *and* $E_{min} \geq 3$.

2. To avoid the additional complication of establishment entry and exit, I use a balanced panel. To have a balanced panel, I face a trade-off between the wide and the longitudinal dimension of the panel. A balanced panel for the entire sample period would leave me with too few establishment observations. I pick a panel length of four years, and then select the four consecutive years for which I would have the maximal number of uninterrupted establishment-year observations.

3. I drop establishments which have more than 300 employees in any of the survey years, thus dropping 17% of the remaining establishments. This reduces the representativeness of the sample but also its size heterogeneity (the maximum number of employees in this sample exceeds 20,000). Presumably the excluded establishments (a long right tail) might have very different adjustment costs.

[8]Let E_{max} be the greater of the two employment records, and E_{min} the smaller, and let $\Delta \equiv E_{max}/E_{min}$. Any observation with $\Delta \geq 2$ is dropped. In addition, I drop observations if $\Delta \geq 1.66$ *and* $E_{max} \geq 20$, and if $\Delta \geq 1.5$ *and* $E_{max} \geq 50$.

4. Even though I observe the wage paid to every individual employee, the total wage bill at time t is $E(w_t)l_t$. Instead of the mean wage I use the median due to some issues of right-censoring (due to caps to the social security contributions).

5. I define a hire h in period t as an employment spell which starts between the first and last day of the period, *and* if the employee has not been employed by the establishment at any point in the previous 366 days.[9] If the employee has been employed in the establishment in the previous 366 days, I treat him as a recalled employee and implicitly assume that this is done without incurring hiring costs. The recalled worker will still be counted in the employment stock.

6. I define a separation f in period t as an employment spell terminating between the last day of period $t-1$ (e.g. the last day of work is the last day of the finished month) and the day before the last day of t, *and* if the employee is not "recalled" within the next 366 days, *and* if the separation is not from a worker aged 60 or older who does not start a new employment spell subsequently. With this last condition I intend to capture retirements, which I also assume do not cause any separation costs.

7. I define employment at time t to be the number of social security paying employees on the last day of period t.

Finally, I use a randomly selected 90% of the establishment for the estimation, leaving the remaining 10% for out-of-sample predictions. These choices are admittedly somewhat arbitrary. But one has to take a stand regarding how to treat data inconsistencies, and what to count as a – potentially costly – hire and separation. I also estimate the model where hires and separations are defined more loosely. I discuss this in the results section. These steps leave me with 2,816 establishments. For monthly adjustment frequencies I obtain 33,408 establishment-month observations (a few establishments have to be dropped if no employee or no wage is reported for a month). I do not claim that this sample is representative in any way of all German establishments. But I repeat that estimating "correct" adjustment costs per se is not the purpose of this paper. Instead I am interested in the effect of different model specifications on adjustment cost estimates.

Figure 1 plots the size (in terms of employment) distribution of the establishments. Most establishments are small with only a few employees, but the right tail is very long. The more interesting information is shown in figure 2, which shows net labor adjustments for monthly and for annual frequencies. Note that the

[9] I also estimate the model with a different set of hiring and separation classifications. See results section.

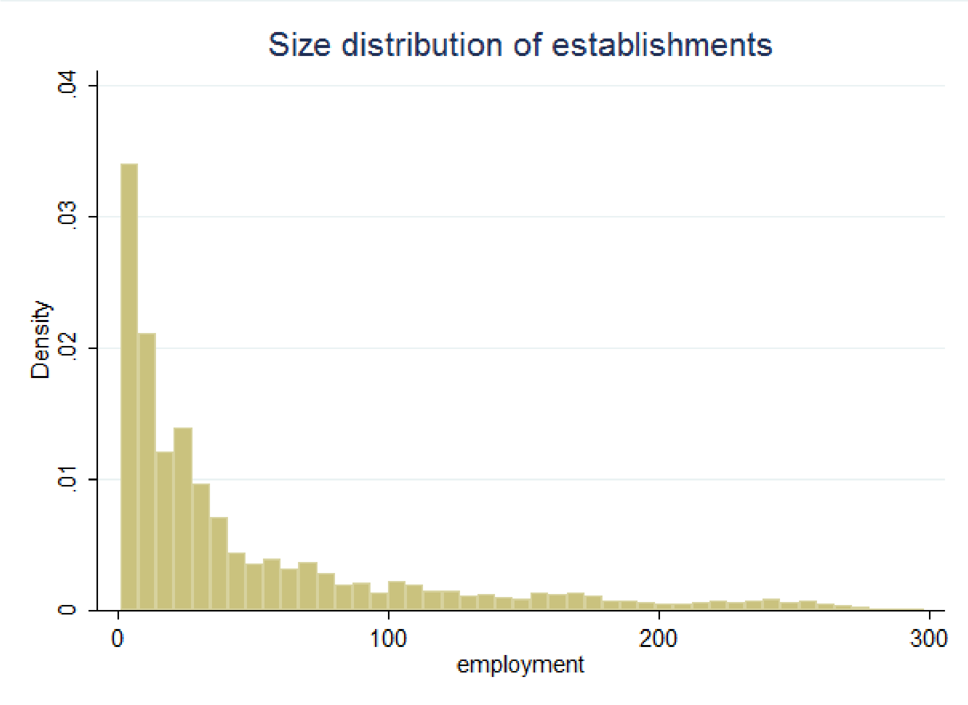

Figure 1: Size distribution

Table 1: Net vs. gross adjustments (percentage of observations)

	Month		Year	
	Net	Gross	Net	Gross
Only hire (H)	19.2	13.2	44.6	7.7
Only separation (F)	19.8	15.2	33.6	6.8
No hire, no separation (P)	61.0	55.6	21.8	10.2
Both hire and separation (B)	n.a.	15.9	n.a.	75.3
Total	100	100	100	100

scales for the two graphs are different. Within a month, about 60% of all establishment-month observations do not change their employment stock (this masks some cases of hires and separations of equal quantity), and almost no establishment has net changes of more than 5 employees. Contrast this with annual adjustments: No adjustment occurs only in little more than 20% of all cases, and the tails are much fatter. There is nothing surprising about this. But the implications for adjustment cost estimates are profound.

Finally, to give a first impression on the importance of distinguishing net from gross adjustments, in table 1 I tabulate the frequencies of each discrete choice for both gross and net changes. We see that even for monthly frequencies net adjustments "hide" many cases (16%) of simultaneous hiring and separations. Importantly, the establishments are not as inactive as one might think observing only net adjustments (55.6% instead of 61%). For annual adjustment, the difference between gross and net adjustment is dramatic. 75.3% of establishments do hire and separate within a year.

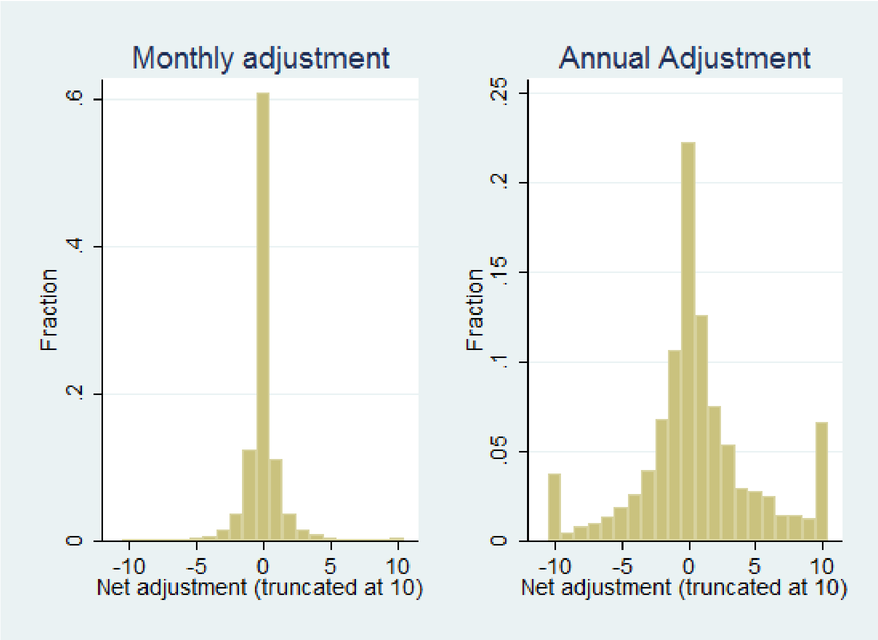

Figure 2: Net adjustment

4.1 Sales

A substantial problem is posed by the fact that the sales variable is taken from the establishment survey and thus only available as an annual variable. All other relevant variables (hires, separations, wage, and employment) can be constructed for any day, but I can not know the revenues for any other time interval other than a calendar year. Since I will estimate my model using monthly, quarterly, and annual frequencies, I have to decide how the reported annual sales should be divided unto months and quarters. I follow two approaches. The first is to assume that sales were evenly distributed for the year. The second is to assume that sales over months and quarters add up to the annual sales number reported by the establishment, that they are never negative, but that changes from period to period are smooth. In particular, taking for example monthly sales, I construct sales to solve the following problem:

$$
\begin{aligned}
\min_{\{s_t\}_{t=1}^{T}} J = \sum_{t=2}^{T} (s_t - s_{t-1})^2 \quad \text{s.t.} \\
\sum_{t=1}^{12} = S_1 \\
\sum_{t=13}^{24} = S_2 \\
\vdots \\
s_t \geq 0 \quad \forall t
\end{aligned}
$$

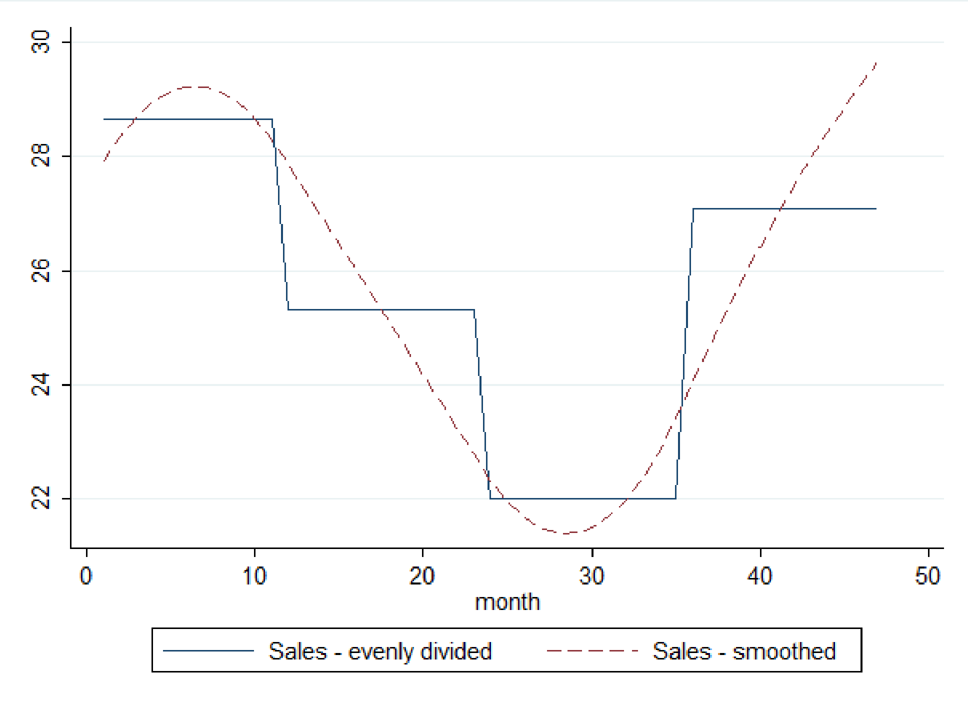

Figure 3: Annual sales divided to months, even and smooth

where S are the reported annual sales. Figure 3 depicts the two sales series for an establishment. Of course both series are "wrong". They contain both classical measurement error due to the establishment probably reporting only a rounded or an approximate sales number, as well as error due to wrongly allocating the annual sales to months and quarters. Moreover, in both of the series there will almost certainly be autocorrelation in the measurement error. But recall that in the estimation the individual sales data are not actually used. Rather, the *average* (over time and establishments) sales conditional on the state and the discrete choice are used, that is $\mathbb{E}(S_t|\mathbf{x}_t, d_t)$, so that the errors in the individual sales data should to some extent cancel out. All sales data are net sales, that is revenue minus expenditures on intermediate inputs. This measurement error also explains the reason for the parameter σ_s in equation 7. Remember that I estimate expected sales $\mathbb{E}(S_t|\mathbf{x}_t)$. Suppose the measurement error a_t is independent and identically distributed, and that measured sales $\tilde{S}$ relate to true sales S as $S_t = a_t\tilde{S}_t$. Then $\mathbb{E}(S_t|\mathbf{x}_t) = \mathbb{E}(a_t\tilde{S}_t|\mathbf{x}_t)$. If the covariance between a_t and $\tilde{S}_t$ is zero, then we have $\mathbb{E}(S_t|\mathbf{x}_t) = \mathbb{E}(a_t\tilde{S}_t|\mathbf{x}_t) = \mathbb{E}(a_t)\mathbb{E}(\tilde{S}_t|\mathbf{x}_t)$. Thus, $\mathbb{E}(a_t)$ is $1/\sigma_s$ in equation 7.

4.2 Capital

The literature on factor adjustment costs has mostly assumed that adjustment of all other factors is costless, and the present paper is no exception. Bloom (2009) is an exception to this simplification and his results suggest that neglecting capital adjustment costs might seriously bias labor adjustment cost estimates. The reason for why this bias would occur is very intuitive. If capital and labor are interdependent, either as substitutes or complements, adjustments of the two factors are likely to be simultaneous. Thus, costs of investing might easily be attributed to labor adjustment. A model including capital adjustment costs would need to include the capital stock as a state variable. The establishment survey includes the value of annual investments (the same difficulty in allocating this annual figure to months and quarters as we saw for sales would apply). Establishments are also asked what percentage of their investments have been used to replace depreciated capital. With external information on depreciation rates (for example industry-wide) an estimate of the capital stock can be constructed. I constructed capital in this way using depreciation rate information from the German Statistical Office, but the thus constructed variable exhibited within-establishment variations which I judged to be not credible (an AR(1) regression of capital produced an R^2 of 0.05). Furthermore, using this variable would expand the state space from three to four variables, thus diminishing cell sizes for the nearest-neighbor estimates outlined above and multiplying estimation times by at least one order of magnitude. All these considerations have led me to not include capital as a separate variable. However, as a “quick and dirty” alternative I have estimated the model with eight instead of four discrete choices, where each original discrete choice is split into *with investment* and *no investment*, and investment incurs a fixed cost. The choice is *with investment* if in the observed calendar year the establishment reported investment of at least 1,000 euros and *no investment* else. By doing this, I intend to capture some of the contamination that might accrue due to simultaneous capital and labor adjustments. This exercise is just intended as an exploration and I make no claim here that this approach solves the problem of interrelated factor demand.

4.3 Hours

The data distinguish between part- and full-time employment. A more detailed variable on hours worked is not available, and I do not use the part- vs. full-time information. Intuitively, the true adjustment costs will be lower than the ones I estimate. We know that establishments can respond to changing circumstances by changing the average hours worked or by changing the number (and/or composition) of employees. We are not observing the instances in which average hours are adjusted. Presumably, if we were able to close the hours adjustment channel (as in my model), we would see more adjustment through hirings and separations in the data. That is, adjustment through number of employees is less inflexible than it appears in my final sample, because establishments can respond to changing circumstances through both channels, of which I observe only one. Without hours adjustment the establishment would be more active in terms of hiring and separations and estimation would yield lower adjustment costs.

5 Results

Recall that the parameter vector to be estimated is $\tilde{\theta}_u = (1/(\sigma_s\sigma) \quad -1/\sigma \quad -\tau^+/\sigma \quad -\tau^-/\sigma)$. It also includes $-\tau^f/\sigma$ if I include a fixed cost of adjustment which is incurred in any period in which hiring or separations take place, or $-\tau^i/\sigma$ if I include a fixed cost of investment as described in the capital section. The main results for monthly adjustment frequencies are presented in table 2. The first two columns are results for models where hiring and separation costs are incurred on NET changes. If an establishment had more hirings than separations I considered its choice to be H, in the reverse case F, and if the two variables were equal the choice is P. The choice B – simultaneous hiring and separations – does not exist in this model. By necessity this would be the model when employment is available only as a stock variable, or if labor is treated as homogeneous, in which case the choice B would be economically non-sensible if adjustment is costly. The first column divides annual sales evenly across the months of the year, and the second shows results where sales have been smoothed according to the procedure described in section 4.1.

Qualitatively the results are in line with expectations. The sale parameter and the scale parameter of the error term are positive, and both hiring and separations are costly. Where the model fails is in delivering adjustment costs in a realistic range. Even though I don't have any priors about these costs,[10] hiring costs of 80,000 Euros per hire and separation costs of 110,000 Euros per separation seem too high.[11] An interesting result is that the results are insensitive to which sales measure I use. This is because individual measurement errors due to dividing annual sales largely cancel each other out – irrespective of how the division is carried

[10]The only estimated hiring costs for German establishments are from cite and lie between 4,000 and 6,000 Euros.

[11]Median annual earnings in the data are around 19,000 Euros.

Table 2: Adjustment cost estimates (in 1,000 euros)

Discrete choices	Three	Three	Four	Four	Four	Eight
Sales	Evenly divided	Smoothed	Evenly divided	Smoothed	Smoothed	Smoothed
$1/(\sigma_s\sigma)$	0.0018	0.0015	0.0022	0.0021	0.0008	0.0023
$-1/\sigma$	-0.0045	-0.0042	-0.0057	-0.0059	-0.0041	-0.0065
$-\tau^+/\sigma$	-0.36	-0.34	-0.23	-0.21	0.22	-0.24
$-\tau^-/\sigma$	-0.49	-0.48	-0.24	-0.23	0.33	-0.25
$-\tau^f/\sigma$					-1.96	
$-\tau^i/\sigma$						-14.18
Hiring cost (τ^+)	80.85	79.41	40.84	35.92	-53.79	36.14
Separation cost (τ^+)	108.73	113.19	41.49	38.88	-80.63	38.03
Fixed cost (τ^+)					481.54	
Investment cost (τ^+)						2164.39

out – when creating average sales conditional on the state and the discrete choice. Columns three and four add the fourth option B to the model. This decreases the frequency of P choices in the data, since in the 3-choice model some cases in which both hirings and separations occur would have been classified as P. In line with the intuition given in the identification section, less inactivity should increase the cost parameters, and this is indeed what we observe. The cost parameters in the 4-choice model are 50 to 65 percent smaller than in the 3-choice model. Moreover, hiring and separations seem equally costly, while in the 3-choice model separations were 30% more costly than hirings. This is a crucial result. It demonstrates that the distinction between net and gross adjustments is very important for the estimated costs. Column 5 adds a fixed – and symmetric – adjustment cost to the model. Unfortunately this model does not work well. The fixed cost of adjustment is extremely high, and every additional hire or separation actually reduces this cost.[12] Finally, the sixth column shows results from a model where the investment choice has been added to the choice set, creating eight discrete choices as described in section 4.2. While investment clearly is an important cost factor, it is surprising that the remaining parameters are only negligibly affected. The adjustment costs are virtually equal to the result of column 4. As mentioned earlier, this last specification is misspecified since capital is not included as a state variable, but the result suggests that investment is orthogonal to the remaining variables of this model.

Table 3 presents results for different assumptions on the frequency of hiring and separation decisions. All models are estimated with four discrete choices. The first two columns are for monthly data and replicate columns 3 and 4 from table 2. The next two columns show results from the model where decisions are made at the beginning of each quarter, and the last column shows results from annual adjustments. Since sales are

[12]All models with fixed adjustment costs have exhibited this result, so I do not pursue this specification any further.

Table 3: Adjustment costs – Different choice frequencies (in 1,000 euros)

Frequency Sales	Monthly Evenly divided	Monthly Smoothed	Quarterly Evenly divided	Quarterly Smoothed	Annually not applicable
$1/(\sigma_s\sigma)$	0.0022	0.0021	0.0003	0.0003	0.0018
$-1/\sigma$	-0.0057	-0.0059	-0.0002	-0.0003	0.0031
$-\tau^+/\sigma$	-0.23	-0.21	0.06	0.06	-0.41
$-\tau^-/\sigma$	-0.24	-0.23	0.10	0.11	-0.54
Hiring cost (τ^+)	40.84	35.92	-300.73	-184.14	-135.39
Separation cost (τ^+)	41.49	38.88	-496.29	-333.92	-175.55
Observations	33,408		11,152		2,816

reported in annual frequency, this variable can be taken as it is reported.

The parameter and cost estimates are highly sensitive to the choice of the assumed adjustment frequency. Comparing monthly to quarterly frequency we can rely on the intuition outlined in section 3.1: Less inactivity is observed in quarterly data. Consequently, the parameters $-\tau^+/\sigma$ and $-\tau^-/\sigma$ are greater than in the monthly model, to the extent that hiring and separations are not costly any more. At quarterly frequency, the data clearly rejects the adjustment cost model. Why this relationship does not hold up for annual adjustments is puzzling, but the reader should bear in mind that the data is reduced to one twelfth of the observations (2,816), considerably increasing the noise in the data.

Consider next in table 4 how results respond to changes in the model assumptions or sample selection. All results are for monthly adjustment freqencies. Since the two sales variables yield very similar results, I only report results for the evenly divided sales variable. I first compare the benchmark model (table 2, column 3) to a model with lower discount factors, one with an annual discount factor β of 0.5, and one with $\beta = 0$, making this a static model, e.g. the firm is not forward looking. Decreasing the discount factor reduces the magnitudes of the sales and wage variables, since the expected future stream is discounted. In a static model, only the differences between alternatives in the current period matter. In a forward-looking model, it is the differences in the current and the expected future flows associated with the choices. This "inflation" of the variables does not occur with the number of hires and separations. We see the outcome of this in the second column: The magnitude of the coefficient $-1/\sigma$ increases because the magnitude of the wage difference between alternatives decreases. Therefore the adjustment cost estimates are reduced. The same argument applies for the static model in column 3. The wage coefficient increases in magnitude. But we also see that the hiring and separation parameters increase in magnitude. An informal look at the data reveals that contemporaneous wages and adjustments are much more correlated than their present discounted

Table 4: Adjustment costs – Alternative specifications (in 1,000 euros)

Specification	Benchmark	$\beta = 0.5$	$\beta = 0.0$	Alternative adjustment	Small firms	Cost heterogeneity
$1/(\sigma_s \sigma)$	0.0022	0.0023	0.0017	0.0019	0.0033	0.0020
$-1/\sigma$	-0.0057	-0.0082	-0.0832	-0.0048	-0.0128	-0.0064
$-\tau^+/\sigma$	-0.23	-0.20	-0.05	-0.21	-0.38	-0.56
$-\tau^-/\sigma$	-0.24	-0.23	-0.19	-0.15	-0.79	0.22
$-\frac{\tau^+ * \log(\bar{w})}{\sigma}$						-0.51
$-\frac{\tau^- * \log(\bar{w})}{\sigma}$						0.18
Hiring cost (τ^+)	40.84	24.93	0.63	42.62	29.83	
Separation cost (τ^+)	41.49	27.47	2.28	31.24	61.88	

values. Hiring and separation costs are clearly much lower when the discount factor is reduced.

Next, I estimated the model defining a hire to be the beginning of an employment spell if the employee has not worked for the same employer in the previous 185 days, and a separation the end of an employment spell if the employee will not work for the same employer within the next 185 days. Importantly, compared to the benchmark model, retirements are now also counted as separations. As this clearly creates more separations, we should observe *lower* separation estimates, and this is indeed what we observe. The hiring cost parameter is hardly affected. Next, I constrain my sample of firms to those with at most 100 employees (compared to 300 in the benchmark model). This gives an indication about the effect of aggregation in the firm size. Smaller firms will adjust less frequently, and the cost parameters should be greater in magnitude. This is what we observe in column 5. Finally, I allowed the adjustment costs to be dependent on the average log wage paid in the establishment by including the interaction terms $\mathbb{E}(h_t|\mathbf{x}_t, d_t \in \{H, B\}) * \log(\bar{w})$ and $\mathbb{E}(f_t|\mathbf{x}_t, d_t \in \{F, B\}) * \log(\bar{w})$ into the $\mathbf{z}(\mathbf{x}_t, d_t)$ vector. The results in column 6 suggest that if average wages in the establishment are higher, adjustment costs are lower, e.g. a 10% increase in wages increases hiring costs by 0.22*0.1 = 0.022 thousand (220) Euros, and increases separation costs by 180 Euros. This is likely to be a consequence of greater establishments paying higher wages.

6 Performance

I now turn to an issue which in my opinion in the framework of structural models has remained somewhat unsatisfactory. How can we evaluate/validate a structural model? And against which benchmark? An ideal setting would be to estimate a structural model for a certain time period, and use the estimated model to predict outcomes for another time period in which one of the structural parameters is changed (I will call this "treatment" in line with the treatment effect and evaluation literature). This is the route followed by Aguirregabiria and Alonso-Borrego (2014). That would be informative about the usefulness of the model. A well-designed, reduced form policy evaluation would give us a credible treatment estimate. Since the unique purpose of the latter is to quantify the treatment effect, the structural modeller could not hope to outperform the reduced form estimate with the predictions of the structural model. Comparing structural models with reduced-form policy evaluations in this way seems misguided, since the purpose of structural models is not primarily, or even if it is, not uniquely, to evaluate a particular treatment. Be that as it may, the present paper does not have a structural change to evaluate.

A sensible evaluation would be based on comparing the goodness of fit (e.g. the Pseudo R-squared based on the likelihood function) and the fraction of correctly predicted outcomes between the structural model and a naive reduced form model. This can be done for the estimation sample, or for out-of-sample observations. What should the naive model be? If I had estimated the structural model based on three outcomes (that is net hiring, net inactivity, and net separations), my model should at least outperform a simple ordinal choice model such as ordinal logit with sales, average wages, and the employment stock (the same state variables as in the structural model) as explanatory variables. But one of the strengths of the present paper is precisely having the fourth choice of both hiring and separations simultaneously, and the four choices do not have a natural ordering. Thus, the best comparison I can make is with respect to a multinomial logit model consisting of the aforementioned four choices, and using the same $\mathbf{z}(\mathbf{x}_t, d_t)$ as in the full dynamic model to characterise the choices, but disregarding the dynamic aspect. This amounts to using $\mathbf{z}$ instead of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{e}}$ in the likelihood function. The test is therefore about whether modelling this problem as a dynamic discrete choice model improves on treating it as static. I present two performance measures: the pseudo R^2 defined as $1 - LL_u/LL_r$, where LL_u is the log-likelihood of the model and LL_r is the log-likelihood of assigning a probability of 25% to each choice; and the fraction of correctly predicted outcomes, where a correct prediction is defined as the observed choice having a higher probability of being chosen than all alternatives. These two measures are calculated for 1) the non-parametric, initial choice probabilities described in the estimation section, 2) the dynamic model, 3) the static model. All measures are calculated for a randomly selected 10% of the original sample which were excluded from the estimation sample.

Table 5 shows the performance measures for these three models. As expected, the non-parametric

Table 5: Out of sample performance

	Pseudo R^2	Fraction predicted
Non-parametric	0.260	0.63
Dynamic Model	0.088	0.40
Static Model ($\beta = 0$)	0.079	0.42

estimates outperform the parametric predictions. The parametric models are based on four parameters, the non-parametric model separately calculates choice probabilities for 8,680 points in the state space. It serves here as an ideal benchmark. The interesting comparison is between the dynamic and the static model. No significant difference exists between the performances of the two models. The pseudo R^2 of the dynamic model is slightly higher, but the fraction correctly predicted even slightly lower than in the static model. Accounting for the dynamic aspects of an establishment adds nothing to the model performance in terms of out-of-sample predictions despite having very different estimated adjustment costs – a phenomenon that relates to the poor identification of the discount parameter in these types of models (see Abbring and Daljord (2016) for a treatment of this). The performances are similar because both models fit the data using the same choice and the same state variables. But the estimates are very different because the differences between the values of the different choices are much greater in the dynamic model than in the static model, since the former is composed of the entire expected future stream of profits associated with a choice. This opens up an interesting question of model selection regarding the dynamic aspects of dynamic discrete choice models.

7 Discussion and conclusion

I have analysed labor adjustment costs using a structural dynamic discrete choice model of establishments' hiring and separation decisions. I did this using linked employer-employee data from Germany which allowed me to observe all the in- and out-flows of employees in an establishment, and thus to distinguish net from gross flows. My objective was to analyse which model specifications – in terms of temporal aggregation and net vs. gross adjustments – yield cost estimates in accordance with economic intuition.

Using monthly choice frequencies, the signs of the four parameters are in line with economic intuition and with model assumptions. Hirings and separations are costly, the scale parameter of the extreme value distribution is positive, sales are profitable, and paying wages is unprofitable (ceteris paribus). One of the main objectives of the paper was to show that the timing assumptions, that is the assumed frequency of revising hirings and separations, should matter for the estimated costs, and I have demonstrated that they indeed matter a great deal. The change in the parameters in moving from monthly to less frequent choice frequencies changed the cost parameters in line with my predictions. Indeed, the estimation results clearly rejected quarterly and annual adjustment frequencies. Another main objective was to highlight the importance between gross and net changes. Again, in line with predictions, I showed that adding the choice of simultaneous hiring and separations result in smaller adjustment costs – the reduction is more than 50%. I also investigated a number of other specifications and sample restrictions and found that adjustment costs are estimated to be higher for smaller establishments, and that less restrictive definitions of separations decrease the cost parameter. All these results work through the channel of how much activity is observed. Shorter choice frequencies, not classifying retirements as separations, using smaller establishments, and using only net adjustments all result in higher adjustment costs. This is intuitive. Economic conditions change all the time. If firms do not respond frequently by adjusting employment it must be costly to do so. In that sense the present analysis has been very instructive.

However, the dynamic model has not outperformed a static one but has resulted in very different adjustment cost quantities (e.g. a separation costs 2,000 Euros in the static, and close to 40,000 Euros in the dynamic model). Given the similar performances of the static and dynamic model and the very different cost estimates, an open question is how one could discriminate between the two models to reject one in favour of the other.

References

Abbring, Jaap H and Øystein Daljord (2016), "Identifying the discount factor in dynamic discrete choice models." *CEPR Discussion Papers*, 11133.

Abel, Andrew B and Janice C Eberly (1994), "A unified model of investment under uncertainty." *American Economic Review*, 84, 1369–1384.

Abowd, John M, Patrick Corbel, and Francis Kramarz (1999), "The entry and exit of workers and the growth of employment: an analysis of French establishments." *Review of Economics and Statistics*, 81, 170–187.

Abowd, John M and Francis Kramarz (2003), "The costs of hiring and separations." *Labour Economics*, 10, 499–530.

Aguirregabiria, Victor (1997), "Estimation of dynamic programming models with censored dependent variables." *Investigaciones Economicas*, 21, 167–208.

Aguirregabiria, Victor (1999), "The dynamics of markups and inventories in retailing firms." *Review of Economic Studies*, 66, 275–308.

Aguirregabiria, Victor and Cesar Alonso-Borrego (2014), "Labor contracts and flexibility: Evidence from a labor market reform in Spain." *Economic inquiry*, 52, 930–957.

Aguirregabiria, Victor and Pedro Mira (2002), "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models." *Econometrica*, 70, 1519–1543.

Aguirregabiria, Victor and Pedro Mira (2010), "Dynamic discrete choice structural models: A survey." *Journal of Econometrics*, 156, 38–67.

Asphjell, Magne K, Wilko Letterie, Øivind A Nilsen, and Gerard A Pfann (2014), "Sequentiality versus simultaneity: Interrelated factor demand." *Review of Economics and Statistics*, 96, 986–998.

Bentolila, Samuel and Giuseppe Bertola (1990), "Firing costs and labour demand: how bad is eurosclerosis?" *Review of Economic Studies*, 57, 381–402.

Bloom, Nicholas (2009), "The impact of uncertainty shocks." *Econometrica*, 77, 623–685.

Burgess, Simon, Julia Lane, and David Stevens (2001), "Churning dynamics: an analysis of hires and separations at the employer level." *Labour Economics*, 8, 1–14.

Cooper, Russell, John Haltiwanger, and Jonathan L Willis (2010), "Euler-equation estimation for discrete choice models: A capital accumulation application." *NBER*, 15675.

Cooper, Russell, John Haltiwanger, and Jonathan L Willis (2015), "Dynamics of labor demand: Evidence from plant-level observations and aggregate implications." *Research in Economics*, 69, 37–50.

Ejarque, João Miguel and Pedro Portugal (2007), "Labor adjustment costs in a panel of establishments: A structural approach." *IZA*, 3091.

Goolsbee, Austan and David B Gross (2000), "Estimating the form of adjustment costs with data on heterogeneous capital goods." *mimeo*.

Hamermesh, Daniel S (1989), "Labor demand and the structure of adjustment costs." *American Economic Review*, 79, 674–689.

Hamermesh, Daniel S (1992), "A general model of dynamic labor lemand." *Review of Economics and Statistics*, 74, 733–737.

Hamermesh, Daniel S and Gerard A Pfann (1996), "Adjustment costs in factor demand." *Journal of Economic Literature*, 34, 1264–1292.

Hopenhayn, Hugo and Richard Rogerson (1993), "Job turnover and policy evaluation: A general equilibrium analysis." *Journal of Political Economy*, 915–938.

Hotz, V Joseph and Robert A Miller (1993), "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies*, 60, 497–529.

Jacobebbinghaus, Peter (2008), "LIAB Datenhandbuch." *FDZ Datenreport*, 3, 2008.

Kramarz, Francis and Marie-Laure Michaud (2010), "The shape of hiring and separation costs in France." *Labour Economics*, 17, 27–37.

Lapatinas, Athanasios (2009), "Labour adjustment costs: Estimation of a dynamic discrete choice model using panel data for Greek manufacturing firms." *Labour Economics*, 16, 521–533.

Mühlemann, Samuel and Harald Pfeifer (2013), "The structure of hiring costs in Germany: Evidence from firm-level data." *IZA*, 7656.

Nilsen, Øivind A, Kjell G Salvanes, and Fabio Schiantarelli (2007), "Employment changes, the structure of adjustment costs, and plant size." *European Economic Review*, 51, 577–598.

Pagan, Adrian and Aman Ullah (1999), *Nonparametric econometrics*. Cambridge university press.

Pfann, Gerard A and Franz C Palm (1993), "Asymmetric adjustment costs in non-linear labour demand models for the Netherlands and UK manufacturing sectors." *Review of Economic Studies*, 60, 397–412.

Rota, Paola (2004), "Estimating labor demand with fixed costs." *International Economic Review*, 45, 25–48.

Rust, John (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica*, 55, 999–1033.

Scarf, Herbert (1959), "The optimality of (S,s) policies in the dynamic inventory problem." In *Mathematical Methods for the Social Sciences* (Kenneth J Arrow, Samuel Karlin, and Patrick Suppes, eds.), chapter 13, 196–202, Stanford University Press, Stanford.

Stokey, Nancy and Robert E Lucas (1989), *Recursive Methods in Economic Dynamics*. Harvard University Press.

Varejão, José and Pedro Portugal (2007), "Employment dynamics and the structure of labor adjustment costs." *Journal of Labor Economics*, 25, 137–165.

Appendix - Derivation of $\mathbf{W}(\mathbf{x})$

Recall the choice-specific value function

$$v(\mathbf{x}_t, d_t) = u(\mathbf{x}_t, d_t) + \beta \mathbb{E}_{\mathbf{x}_{t+1}, \varepsilon_{t+1}} \left(V(\mathbf{x}_{t+1}, \varepsilon_{t+1}) | \mathbf{x}_t, d_t \right)$$

I can rewrite this as

$$v(\mathbf{x}_t, d_t) = u(\mathbf{x}_t, d_t) + \sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, \varepsilon_{t+i}} \left(u(\mathbf{x}_{t+i}, d^*_{t+i}) + \varepsilon^{d^*_{t+i}} | \mathbf{x}_t, d_t \right) \tag{9}$$

I can separately characterize the parts containing u and ε. For u:

$$\begin{aligned}
& u(\mathbf{x}_t, d_t) + \sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, \varepsilon_{t+i}} \left(u(\mathbf{x}_{t+i}, d^*_{t+i}) | \mathbf{x}_t, d_t \right) \\
&= u(\mathbf{x}_t, d_t) + \sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, | d_t, \mathbf{x}_t} \left[\sum_{d_{t+i} \in D} P(d^*_{t+i} = d_{t+i} | \mathbf{x}_{t+i}) u(\mathbf{x}_{t+i}, d_{t+i}) \right] \\
&= \mathbf{z}(\mathbf{x}_t, d_t) \theta'_u + \sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, | d_t, \mathbf{x}_t} \left[\sum_{d_{t+i} \in D} P(d^*_{t+i} = d_{t+i} | \mathbf{x}_{t+i}) \mathbf{z}(\mathbf{x}_{t+i}, d_{t+i}) \theta'_u \right] \\
&= \left(\mathbf{z}(\mathbf{x}_t, d_t) + \sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, | d_t, \mathbf{x}_t} \left[\sum_{d_{t+i} \in D} P(d^*_{t+i} = d_{t+i} | \mathbf{x}_{t+i}) \mathbf{z}(\mathbf{x}_{t+i}, d_{t+i}) \right] \right) \theta'_u
\end{aligned} \tag{10}$$

For ε:

$$\sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i}, \varepsilon_{t+i}} \left(\varepsilon^{d^*_{t+i}} | \mathbf{x}_t, d_t \right)$$

$$\sum_{i=1}^{\infty} \beta^i \mathbb{E}_{\mathbf{x}_{t+i} | \mathbf{x}_t, d_t} \left[\sum_{d_{t+i} \in D} P(d^*_{t+i} = d_{t+i} | \mathbf{x}_{t+i}) \mathbb{E} \left(\varepsilon^{d_{t+i}} | \mathbf{x}_{t+i}, d^*_{t+i} = d_{t+i} \right) \right] \tag{11}$$

The expression $\mathbb{E} \left(\varepsilon^{d_{t+i}} | \mathbf{x}_{t+i}, d^*_{t+i} = d_{t+i} \right)$ is the expected value of the shock to choice d_{t+i} given that this choice dominated all alternatives. The important insight of Hotz and Miller (1993) is to show that this can be expressed as a function of the choice probabilities. In the case of the type I extreme value distribution, this expectation is given by[13]

$$\mu+\sigma(\gamma-\ln P(d_t|x_t))\equiv \mathbf{e}(\mathbf{x}_t,d_t)\theta_e'$$

with $\mathbf{e}(\mathbf{x}_t,d_t)=(1 \quad \gamma-\ln P(d_t|\mathbf{x}_t))$, $\theta_e=(\mu \quad \sigma)$, and γ is Euler's constant. The parameters of the distribution (μ and σ) are usually not identified and set to 0 and 1 respectively. Replacing this in equation 11 I get

$$\begin{aligned}&\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}|\mathbf{x}_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}_{t+i})\mathbf{e}(\mathbf{x}_t,d_t)\theta_e'\right]\times\\ &\left(\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}|\mathbf{x}_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}_{t+i})\mathbf{e}(\mathbf{x}_t,d_t)\right]\right)\theta_e' \qquad (12)\end{aligned}$$

Now, having derived expressions 10 and 12, I rewrite 9 as

$$v(\mathbf{x}_t,d_t)=\tilde{\mathbf{z}}(\mathbf{x}_t,d_t)\theta_u'+\tilde{\mathbf{e}}(\mathbf{x}_t,d_t)\theta_e' \qquad (13)$$

where

$$\tilde{\mathbf{z}}(\mathbf{x}_t,d_t)=\mathbf{z}(\mathbf{x}_t,d_t)+\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i},|\mathbf{x}_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}_{t+i})\mathbf{z}(\mathbf{x}_{t+i},d_{t+i})\right] \qquad (14)$$

and

$$\tilde{\mathbf{e}}(\mathbf{x}_t,d_t)=\sum_{i=1}^{\infty}\beta^i\mathbb{E}_{\mathbf{x}_{t+i}|\mathbf{x}_t,d_t}\left[\sum_{d_{t+i}\in D}P(d^*_{t+i}=d_{t+i}|\mathbf{x}_{t+i})\mathbf{e}(\mathbf{x}_t,d_t)\right] \qquad (15)$$

If $\mathbf{x}$ is discrete with distribution function $f(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)$, following Aguirregabiria and Mira (2010), $\tilde{\mathbf{z}}(\mathbf{x}_t,d_t)$ and $\tilde{\mathbf{e}}(\mathbf{x}_t,d_t)$ can be shown to have the following recursive expressions:[14]

$$\tilde{\mathbf{z}}(\mathbf{x}_t,d_t)=\mathbf{z}(\mathbf{x}_t,d_t)+\beta\sum_{\mathbf{x}_{t+1}}f(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}_{t+1})\tilde{\mathbf{z}}(\mathbf{x}_{t+1},d_{t+1})\right)$$

$$\tilde{\mathbf{e}}(\mathbf{x}_t,d_t)=\beta\sum_{\mathbf{x}_{t+1}}f(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D}P(d_{t+1}|\mathbf{x}_{t+1})\left(\mathbf{e}(\mathbf{x}_{t+1},d_{t+1})+\tilde{\mathbf{e}}(\mathbf{x}_{t+1},d_{t+1})\right)\right)$$

[13] I am greatly thankful to Aureo de Paula for providing me with a reference with the proof of this.

[14] The derivation is straightforward and in the next section of the appendix.

Now define

$$\mathbf{W_z}(\mathbf{x}) = \sum_{d \in D} P(d|\mathbf{x})\tilde{\mathbf{z}}(\mathbf{x},d)$$

$$\mathbf{W_e}(\mathbf{x}) = \sum_{d \in D} P(d|\mathbf{x})\left(\mathbf{e}(\mathbf{x},d) + \tilde{\mathbf{e}}(\mathbf{x},d)\right)$$

so that

$$\tilde{\mathbf{z}}(\mathbf{x},d) = \mathbf{z}(\mathbf{x},d) + \beta \sum_{\mathbf{x}'} f(\mathbf{x}'|\mathbf{x},d)\mathbf{W_z}(\mathbf{x}')$$

$$\tilde{\mathbf{e}}(\mathbf{x},d) = \beta \sum_{\mathbf{x}'} f(\mathbf{x}'|\mathbf{x},d)\mathbf{W_e}(\mathbf{x}')$$

where $\mathbf{x}'$ denotes the state vector in the following period. Writing $\mathbf{W}(\mathbf{x}) \equiv [\mathbf{W_z}(\mathbf{x})\mathbf{W_e}(\mathbf{x})]$, $\mathbf{W}(\mathbf{x})^*$ is the unique solution to the recursive equation

$$\mathbf{W}(\mathbf{x}) = \sum_{d \in D} P(d|\mathbf{x}) \times \left([\mathbf{z}(\mathbf{x},d), \mathbf{e}(\mathbf{x},d)] + \beta F(\mathbf{x}'|\mathbf{x},d)W(\mathbf{x}')\right) \tag{16}$$

Appendix - Derivation of $\tilde{\mathbf{z}}(\mathbf{x}_t, d_t)$ and $\tilde{\mathbf{e}}(\mathbf{x}_t, d_t)$

Derivation of $\tilde{\mathbf{z}}(\mathbf{x}_t, d_t)$:

$$\begin{aligned}
\tilde{\mathbf{z}}(\mathbf{x}_t,d_t) =& \mathbf{z}(\mathbf{x}_t,d_t) + \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1})\mathbf{z}(\mathbf{x}_{t+1},d_{t+1})\right) + \\
& \beta^2 \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1})\left[\sum_{\mathbf{x}_{t+2}} F(\mathbf{x}_{t+2}|\mathbf{x}_{t+1},d_{t+1})\left\{\sum_{d_{t+2}\in D} P(d_{t+2}|\mathbf{x}_{t+2})\mathbf{z}(\mathbf{x}_{t+2},d_{t+2})\right\}\right]\right) \\
& + \cdots \\
=& \mathbf{z}(\mathbf{x}_t,d_t) + \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1})\right. \\
& \left.\left[\mathbf{z}(\mathbf{x}_{t+1},d_{t+1}) + \sum_{k=1}^{\infty} \beta^k \mathbb{E}_{\mathbf{x}_{t+1+k}|\mathbf{x}_{t+1},d_{t+1}}\left\{\sum_{d_{t+1+k}\in D} P(d_{t+1+k}|\mathbf{x}_{t+1+k})\mathbf{z}(\mathbf{x}_{t+1+k},d_{t+1+k})\right\}\right]\right) \\
=& \mathbf{z}(\mathbf{x}_t,d_t) + \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t,d_t)\left(\sum_{d_{t+1}\in D} P(d_{t+1}|\mathbf{x}_{t+1})\tilde{\mathbf{z}}(\mathbf{x}_{t+1},d_{t+1})\right)
\end{aligned}$$

where the last equation follows from

$$\tilde{\mathbf{z}}(\mathbf{x}_t, d_t) = \mathbf{z}(\mathbf{x}_t, d_t) + \sum_{j=1}^{\infty} \beta^j \mathbb{E}_{\mathbf{x}_{t+j}|\mathbf{x}_t, d_t} \left(\sum_{d_{t+j} \in D} P(d_{t+j}|\mathbf{x}_{t+j}) \mathbf{z}(\mathbf{x}_{t+j}, d_{t+j}) \right)$$

Derivation of $\tilde{\mathbf{e}}(\mathbf{x}_t, d_t)$:

$$\begin{aligned}
\tilde{\mathbf{e}}(\mathbf{x}_t, d_t) =& \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t, d_t) \left(\sum_{d_{t+1} \in D} P(d_{t+1}|\mathbf{x}_{t+1}) \mathbf{e}(\mathbf{x}_{t+1}, d_{t+1}) \right) + \\
& \beta^2 \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t, d_t) \left(\sum_{d_{t+1} \in D} P(d_{t+1}|\mathbf{x}_{t+1}) \left[\sum_{\mathbf{x}_{t+2}} F(\mathbf{x}_{t+2}|\mathbf{x}_{t+1}, d_{t+1}) \left\{ \sum_{d_{t+2} \in D} P(d_{t+2}|\mathbf{x}_{t+2}) \mathbf{e}(\mathbf{x}_{t+2}, d_{t+2}) \right\} \right] \right) \\
& + \cdots \\
=& \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t, d_t) \left(\sum_{d_{t+1} \in D} P(d_{t+1}|\mathbf{x}_{t+1}) \right. \\
& \left. \left[\mathbf{e}(\mathbf{x}_{t+1}, d_{t+1}) + \sum_{k=1}^{\infty} \beta^k \mathbb{E}_{\mathbf{x}_{t+1+k}|\mathbf{x}_{t+1}, d_{t+1}} \left\{ \sum_{d_{t+1+k} \in D} P(d_{t+1+k}|\mathbf{x}_{t+1+k}) \mathbf{e}(\mathbf{x}_{t+1+k}, d_{t+1+k}) \right\} \right] \right) \\
=& \beta \sum_{\mathbf{x}_{t+1}} F(\mathbf{x}_{t+1}|\mathbf{x}_t, d_t) \left(\sum_{d_{t+1} \in D} P(d_{t+1}|\mathbf{x}_{t+1}) \left\{ \mathbf{e}(\mathbf{x}_{t+1}, d_{t+1}) + \tilde{\mathbf{e}}(\mathbf{x}_{t+1}, d_{t+1}) \right\} \right)
\end{aligned}$$

where the last equation follows from

$$\tilde{\mathbf{e}}(\mathbf{x}_t, d_t) = \sum_{j=1}^{\infty} \beta^j \mathbb{E}_{\mathbf{x}_{t+j}|\mathbf{x}_t, d_t} \left(\sum_{d_{t+j} \in D} P(d_{t+j}|\mathbf{x}_{t+j}) \mathbf{e}(\mathbf{x}_{t+j}, d_{t+j}) \right)$$