Non-Standard Behaviour in Organisational Economics and Individual Choice

Econometric Evidence from the Field

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2013

vorgelegt von

Thomas Kolaska

Referent:	Professor Dr. Florian Englmaier
Korreferent:	Professor Dr. Uwe Sunde
Promotionsabschlussberatung:	6. November 2013

Danksagung

Vier Jahre sind vergangen, seit ich im Oktober 2009 begann an der “Ludwig-Maximilians-Universität München” zu promovieren. Die Dissertation nun in Händen zu halten, verdanke ich einer Reihe an Personen, ohne deren Unterstützung die Promotion in dieser Form nicht möglich gewesen wäre.

An vorderster Stelle steht hierbei mein Erstbetreuer Florian Englmaier. Bereits weit vor dem Beginn meiner Promotion, ich war damals Master Student an der “Universitat Pompeu Fabra”, wurde ich von ihm – glücklicherweise erfolgreich – überzeugt, an die “Universität München” zurückzukehren. Dort angekommen, wurde ich ungeachtet seiner zeitweisen Berufungen nach Konstanz und Würzburg mit einer nicht für möglich gehaltenen Intensität betreut, welche schließlich unter anderem in einem gemeinsamen Papier – zusammen mit Steve Leider, dessen Kommentare mich sehr weiterbrachten – resultierte. Aber vor allem der Zuspruch in schweren Phasen, in Zeiten in denen ich glaubte, mich in Sackgassen zu befinden, bedeuteten für mich, dass ich nie – während all der vier Jahren nicht – in jenes “Loch” fiel, vor dem sich viele Doktoranden berechtigterweise fürchten. Immer dann Zuversicht zu vermitteln, wenn sie nötig war, war rückblickend mindestens ebenso wichtig wie seine fachliche Unterstützung. Ich glaube nicht, dass viele Doktoranden das Glück einer derartigen Betreuung haben.

Uwe Sunde als Zweitgutachter fragte ich leider erst an, als ich bereits weit fortgeschritten war in meiner Arbeit. Trotzdem profitierten meine Papiere durch die Einarbeitung seiner Kommentare und Anregungen maßgeblich. Besonders in Erinnerung bleibt mir hier ein sehr langes Gespräch mit ihm an einem trüben Winternachmittag, bei dem ich sein ehrliches Interesse an meiner Arbeit spürte.

Ebenso danken möchte ich hier meinem Drittbetreuer Ray Rees. Indirekt habe ich es mitunter ihm zu verdanken, mich zu einer Promotion entschieden zu haben. In einem von ihm angebotenen Seminar, an dem ich als Bachelorstundent teilnahm, entdeckte ich meine Leidenschaft für die Wissenschaft, eine Erkenntnis, die entscheidend meinen Lebensweg beeinflusste. Somit freute es mich besonders, als Ray Rees einwilligte, als Drittbetreuer zu fungieren. In gewisser Weise schloss sich dadurch ein Kreis.

Eine unglaublich schöne und gleichsam arbeitsintensive Zeit hatte ich mit Lukas Buchheim, dessen unermüdliches Interesse und dessen Wunsch zur Perfektion mir nicht gedachte Leistungen abverlangten. Gerade das Projekt mit ihm gilt für mich als Musterbeispiel, wie inspirierend es sein kann, gemeinsam über Probleme nachzudenken.

Sascha Becker ermöglichte es mir als Gastwissenschaftler einige Monate an der "University of Warwick" zu verbringen. Diese Zeit ist mir wärmstens in Erinnerung. In einem äußerst angenehmen Umfeld von Doktorandinnen und Doktoranden, sowie spannenden Seminaren war dies eine der produktivsten Phasen meiner Dissertation. Im Endspurt schrieb ich dort mein drittes Papier.

Danken möchte ich in gleichem Maße meiner Familie, die mir immer unter die Arme griff und versuchte mich zu entlasten. Besonderer Dank gilt Isabelle, die mir in all den Jahren den Rücken frei hielt, mich in arbeitsreichen Phasen besonders unterstützte und klaglos "ertrug", wenn ich mal wieder über Wochen hinweg ein Vielfaches an Zeit mit meinen Koautoren verbrachte und daher nur selten zu Hause war. In gleichem Maße allerdings will ich ihr für die kritische Auseinandersetzung mit meiner Arbeit danken, die mich immer wieder veranlasste, die großen Fragen zu stellen.

Contents

List of Tables

List of Figures

Preface

This dissertation consists of three contributions which analyse non-standard behaviour in organisational economics and individual decision making. Every chapter corresponds to one essay and can be read independently of each other.

Non-standard behaviour of individuals in economic situations has been documented over the last two decades in various laboratory and increasingly many field experiments.[1] Various scholars acknowledge that a non-negligible fraction of individuals seems to systematically deviate from rational behaviour as predicted under standard assumptions – a finding which applies to several dimensions of human behaviour. Two important areas of behavioural research, preferences for fairness and erroneous predictions of future utility, play a central role in this dissertation.

Despite several advances over the last years, the fundamental question of how to evaluate the impact of behavioural economics on other fields of economics remains: Should we amend our standard models in favour of more realism at the potential expense of providing clear-cut statements and policy implications? Indirectly Camerer et al. (2011) agree on that question and propose to "gradually replace simplified models based on stricter rationality" (p. 42) with behavioural models and to view results under standard assumptions as a special case of a "more general, behaviorally grounded theory" (p. 42). Following this approach, however, it is inevitable to develop an understanding for the relevance of any of the behavioural biases which were established in the laboratory.

A first, tentative avenue to qualify the importance of various biases is to move away from the laboratory and test whether behaviour observed under artificial conditions can also be identified in market situations. This proceeding can provide evidence on the robustness and the quantitative importance of non-standard behaviour. Results may act as a guideline to gain confidence whether behavioural aspects can "survive markets" and may even influence general equilibrium outcomes.

This methodological approach will be applied in the following thesis: The focus of each essay lies in the exploitation of field data, implying that empirical evidence presented

[1] See for early laboratory contributions for example Güth et al. (1982) or Thaler (1988).

in this dissertation entirely relies on real-world observations. This comes along with econometric challenges because the scope to control for external influences using field data is truly limited. The lack of control is addressed by exploiting exogenous variation, as done in the first chapter. Here we provide evidence for the existence of erroneous individual purchase decisions under risk.

However, as already stressed by Stefano DellaVigna, the natural step after documenting non-standards behaviour is to identify “how [do] markets and institutions respond to these nonstandard features” (DellaVigna (2009), p. 361). Shedding light on that question is the aim of the second part of this dissertation which combines two contributions in the field of personnel economics, exploring the importance of social preferences in intra-firm interactions between employers and employees.

The first chapter of this dissertation, “Projection Bias under Risk”, is joint effort with Lukas Buchheim. In this contribution, we empirically show that purchase decisions may systematically deviate from predictions based on expected utility theory. Standard models commonly assume that individuals combine all relevant information optimally when deciding under risk. Contrary to this, Loewenstein et al. (2003) suggest that decision makers excessively weigh utility at the current state of the world. In its simplest version, projection bias assumes that individuals predict future utility by combining utility at the current state with predictions based on expected utility theory, even though any informativeness of the current state should already be included in rational expectations. Depending on the predictive power of the current state for future states, such behaviour may seriously bias outcomes.

We test for projection bias in the case of an outdoor movie theatre and find that weather conditions at the purchase day largely drive the number of advance ticket sales. Decisive for the utility, customers derive from a visit at the cinema, however, is weather at the day of the screening because moviegoers are fully exposed to unpleasant weather like rainfall or low temperatures. As weather conditions are highly unstable in the region of the respective open-air cinema and tickets are non-refundable, buying tickets in advance *and* conditioning decisions on the current weather involves (unnecessary) high risk. This observation is consistent with projection bias as both the weather forecast, i.e. the piece of information customers under rational expectations make use of, and current weather are predictive for ticket sales. Furthermore we find that customers seem to be unaware of their erroneous behaviour and do not learn from mistakes. We conclude that from the fact that buying behaviour is independent of the time horizon between purchase date and screening day and likewise is unrelated to customers’ previous experience with the cinema.

Alternative explanations do not have any bite: We show that current weather is a very poor predictor for future weather and – even more important – does not contain information beyond the weather forecast. Furthermore we can rule out that correlations between the number of potential customers and weather are the (sole) driver of the results. Finally, capacity constraints of the cinema are not suitable to explain the observed pattern in advance sales.

In a broader context, the finding can have important implications for markets with higher trade volumes. Behavioural finance literature recently provided several examples in which purchasing decisions of investors are driven by current and past returns of the respective assets.[2] In situations, in which current experience only poorly predicts future returns this behaviour may seriously bias investment and saving decisions with potentially severe consequences for individuals. By providing causal evidence for projection bias, we suggest a behavioural mechanism that is consistent with findings in this literature. Our results may hence help to develop de-biasing strategies, supporting individuals to make more informed decisions in the future.

The second and the third chapter of this thesis are closely related as both contributions make use of the same dataset of British firms, the 5th wave of the "Workplace Employment Relations Survey". More importantly, however, both essays aim to provide answers on how non-standard behaviour, more precisely social preferences of employees, shape labour relations.

The second essay is joint work with Florian Englmaier and Stephen Leider. In this contribution, we relate various organisational choices of firms to the human resource practise of compulsory personality tests for job candidates, which we interpret as a predictor of reciprocity within organisations. Applying theoretical models on moral hazard with reciprocity, we regard reciprocity similarly to gift-exchange behaviour, where reciprocal behaviour of employees needs to be "activated" by a kind gesture of the employer. Most easily this can be achieved with wages which are higher than the employer expects to receive in comparable labour relations, but any other non-pecuniary benefit may trigger the same behaviour. As a response to that gesture, employees may return the gift by increasing effort.

But not all employees may be inclined towards reciprocal incentives which forces firms to select the "right" employees if they want to make use of social preferences: In that respects are personality tests among the most popular screening devices. Such tests are commonly designed on basis of the Five Factor model and aim to uncover personal and

[2] See for example Benartzi (2001), Kaustia and Knüpfer (2008), Barber et al. (2009) or Choi et al. (2009).

social traits of individuals. Some of these traits precisely measure behaviour which is described by reciprocity in laboratory experiments. For that reason, we interpret the use of personality tests for job candidates as an indicator for reciprocity within a firm: Establishments which screen for personality should on average employ workers, who have a higher sensitivity towards gift-exchange motives than workers in firms which do not screen.

We find that personality tests are highly predictive of a series of benefits for employees: Workers in firms with personality tests on average are less likely to receive very low wages and are more likely to be granted non-pecuniary benefits like an employer pension scheme and extended annual leave. Furthermore are these employees eligible to more training and also the matters, the training covers are defined broader, implying a higher general value to the employee. There also seems to be a tendency towards more job security in these firms. It is important to understand that some of these policies require substantial (non-monetary) investments into an employee which partly deprives the employer of the option to discipline the worker later on. If however employees are inclined towards reciprocity and such policies are perceived as a kind gesture, then adverse behaviour against the employer should not occur.

For reciprocity to be a consistent explanation of the previous findings, we conjecture that employers with and without compulsory personality tests should furthermore differ in dimensions which are favourable for the firm and attainable when using reciprocal motivation devices – if employers do not receive benefits from kind behaviour, such policies are unlikely to be cost-efficient. In line with this argument, we find that firms using personality tests make use of team-working more frequently and report to be more successful with regard to financial performance, labour productivity, and provide better product quality. Ability tests are, contrary to personality tests, unable to predict this pattern of benefits for the firm and/or benefits for employees. This strengthens the view that screening for personality is elementary to explain favourable labour relations between employer and employee as opposed to screening per se.

The third chapter intends to shed light on human resource management practises and labour productivity, addressing the fundamental trade-off between granting high levels of discretion to employees on the one hand and effort provision on the other hand. Results from previous studies indeed suggest that autonomy in the workplace and less strictly defined work flows can substantially increase productivity if employers manage to prevent agents from exploiting discretion by high levels of shirking.[3] Providing monetary incentives to employees is the traditional solution of economists, but recently increasingly many

[3]See Ichniowski and Shaw (2003) for a survey.

contributions refer to behavioural mechanisms in order to understand employer-employee relationships. The third chapter of this dissertation likewise focuses on reciprocity as one potential force which can mitigate shirking in labour relations.

Information on three dimensions of human resource management policies – the level of discretion employees enjoy, the generosity of wage payments, and the use of personality tests upon hiring, which I again use as a proxy for the inclination of the average employee within an establishment towards reciprocity – allow me to explore effectiveness of each of these policies first in an isolated context and second when multiple practises are applied simultaneously. In line with previous results, I find that providing high levels of discretion does not have detrimental effects on labour productivity, even if no complementary human resource policies are introduced. This could suggest that against common sense employees do not (immediately) exploit reductions in monitoring.[4] However, if firms in addition to granting substantial autonomy also pay high income to their employees *and* workers were screened for personality when entering the firm, then these firms report to be exceptionally successful with regard to labour productivity. A pattern of paying high wages ("gift") to the "right" employees (personality tests) rendering discretion to become desirable is highly consistent with reciprocal behaviour between employer and employee. This is remarkable, because firms which are similarly structured but do not screen for personality tests report significantly lower labour productivity. The same negative result applies to firms which only screen their applicants for ability instead of personality: A combination of high wages, discretion and competency tests fail to predict high labour productivity implying that not screening itself but screening for personality is crucial when firms want to make use of motivation through social preferences.

These findings match closely with results from a laboratory experiment by Bartling et al. (2012a). In this study, the authors provide evidence for the emergence of highly paid jobs with high degrees of discretion *if* employers have the opportunity to screen job candidates for previous effort provision. The parallelism of the results is remarkable because the results of Bartling et al. (2012a) were generated under artificial conditions in the laboratory whereas these findings use field data, which intrinsically do not allow for the same degree of controlling the environment. In addition to laboratory results, however, I can furthermore provide evidence for the importance of personality tests to understand complementarities in human resource management practises.

The logical next step on this agenda is to aim to causally identify reciprocity as an underlying mechanism in employer-employee relations for firms with personality tests. But even in the light of the findings in this dissertation, it seems plausible to assume that

[4] Cf. Nagin et al. (2002)

social preference can shape labour relations which suggest that markets respond to non-standard behaviour of individuals. Findings in this dissertation are encouraging news with regard to the effect of social preferences on labour market outcomes: The labour force is offered more favourable contracts with high compensation and further benefits and employer may make use of additional channels to motivate their employees. Nevertheless it has to be explored in future research whether the utilisation of social preferences in labour market relations throughout leads to favourable outcomes for all parties involved.

Chapter 1

Projection Bias under Risk[0]

1.1 Introduction

In many economic decision problems, the utility from choice materialises in the future such that individuals need to predict their future utility in order to make informed decisions. While standard economic models assume that individuals predict their utility correctly, Loewenstein et al. (2003) argue that individuals make systematic errors; specifically, people tend to underestimate the extent to which changes in the state of the world alter utility. Hence, predicted future utility (at unknown future states) is biased towards utility at today's state. Loewenstein et al. call this error *projection bias*.

There is accumulating evidence that projection bias affects economic decisions like house, car, and apparel purchases (Conlin et al., 2007; Busse et al., 2012) or college choice (Simonsohn, 2010). For example, in parallel work Busse et al. demonstrate that sales of 4-wheel drive vehicles increase by 6 percent after a snowstorm, that is at times when the weather-related utility from owning a 4-wheel drive is very high. All these papers have in common, however, that the weather-related dimension of utility, which serves as the testing ground for projection bias, is most likely not of primary importance for decision makers. Yet, if this is the case and individuals devote only limited attention to predicting the weather-related dimension of utility for available alternatives, they may be more prone to make errors in that dimension.[1] Therefore, it remains an open question whether people are able to overcome projection bias when their attention is drawn to the state-dependent nature of utility.

In this paper, we test for projection bias in a situation where state-dependent utility is

[0] This chapter is based on joint work with Lukas Buchheim.

[1] See Schwartzstein (2012) for theoretical and Hanna et al. (2012) for empirical evidence on how limited attention may lead to errors.

expected to be at the centre of decision maker's attention: We study online advance sales for an outdoor movie theatre. In this context, projection bias predicts that good weather on the purchase-date leads customers to overvalue their utility from visiting the theatre on the movie-date. Hence, the number of advance sales for the theatre should increase if purchase-date weather is good.

The state-dependent nature of utility in this setting is salient for a number of reasons. First, the presence of risk when buying tickets in advance highlights the possibility of facing a different state of the world in the future. Customers face risk when deciding whether or not to buy tickets today for an outdoor movie night in the future because only movie-date weather – as opposed to purchase-date weather – affects utility. This risk is obvious because tickets are only valid for one particular show and are non-refundable (tickets are a perishable good). Additionally, the ticketing website points out the risk in a clear way by stating: "The show is going to take place regardless of weather conditions. (...) You have to pay for your tickets even if you do not collect them."[2] We demonstrate that there is considerable risk because weather at the location of the theatre is highly variable.

In addition to risk, two further characteristics of the decision problem are expected to de-bias potential customers. First, the weather-related dimension of utility is a very important component of total utility derived from the movie night – in a survey, the majority of customers states that weather is at least as important as the movie shown.[3] Since weather is important to customers, they should devote a considerable amount of attention to predicting weather-related utility correctly. Second, when considering to purchase tickets few days in advance, customers can condition their decision on reliable, unbiased, and free information provided by weather forecasts to overcome projection bias.

Contrary to our conjecture that potential customers are by and large de-biased, we find that variations in purchase-date weather explain variations in advance sales to a large degree, controlling for the weather forecast. Across different time horizons – the number of days tickets bought in advance ranging from one day up to three weeks – a one standard deviation increase in sunshine duration leads to an increase in sales between 10 and 25 percent on average. Our findings are robust to considering different subsets of customers. Notably, the results do not change when we consider the behaviour of customers with prior bad experiences defined as rainfall during a previous show they purchased tickets for. The dependence of ticket orders on current weather is thus prevalent for customers, who had the possibility to learn from previous mistakes.

[2] Authors' translation from https://www.didax.de/kms/index.php [4 October 2012].

[3] We conducted a survey at the theatre on a total of 13 nights, interviewing 443 customers. For details, see Section 1.3.3.

We rule out a number of alternative explanations for this finding. First, we show that purchase-date weather has at most negligible predictive power for movie-date weather, ruling out the possibility that current weather is an informative signal for future weather.

Second, we investigate whether the positive effect of purchase-date weather on aggregate sales merely reflects an increase in the number of potential customers who consider visiting the theatre as an attractive leisure activity without affecting individual decisions directly. This may be the case, for example, if good current weather reminds people of the possibility to visit the theatre. We use a strategy similar to Conlin et al. (2007) to distinguish between the latter explanation and projection bias by looking at the decision to collect the tickets that have been purchased in advance. If, on the one hand, projection bias affects individual purchase decisions, utility of customers is upward biased at times of good purchase-date weather. Then, tickets are mistakenly purchased with a higher likelihood. We therefore expect that the likelihood that customers let their tickets expire increases with better purchase-date weather if projection bias affects decisions. If, on the other hand, individual decisions are unbiased and purchase-date weather solely affects the aggregate number of potential customers, there should be no effect on tickets collected. We find a negative effect of purchase-date weather on the probability that tickets are collected, providing further evidence for projection bias.

Third, we argue that weather-related market interactions cannot explain why sales depend on purchase-date weather. In particular, there may be a "precautionary" rationale for purchasing tickets at times of good weather as the latter may increase the perceived probability for the theatre selling out in advance. However, this seems unlikely to be the sole explanation for our findings for two reasons. First, sales well in advance of the movie-date – when the probability for the theatre to sell out is essentially zero – are also weather-dependent. And second, we show that hourly variations in weather explain hourly changes in ticket sales. This is in line with projection bias but does not fit an explanation based on "precautionary" purchasing motives because the perceived probability of the theatre selling out is unlikely to vary with hourly changes in weather.

By showing that projection bias affects individual decisions even in situations in which the state-dependence of utility is particularly salient, our paper complements the emerging literature on projection bias in economics discussed above. In addition to this literature in economics, there is a number of papers in psychology providing evidence for projection bias. This literature deals mostly with how current visceral states – for example hunger or sexual arousal – affect decision making.[4] See Loewenstein and Schkade (1999) for an overview.

[4]See for example Loewenstein (1996), Loewenstein et al. (1997), Read and van Leeuwen (1998), van Boven and Loewenstein (2003), and Nordgren et al. (2007).

Furthermore, it is important to note that projection bias is observationally equivalent to agents holding subjective beliefs that assess the current state of the world to be more likely in the future (this has been pointed out by DellaVigna, 2009).[5] There is some evidence for agents holding these types of beliefs, which Fuster et al. (2010) call "extrapolation bias". For example, several papers in behavioural finance find that individuals tend to choose assets with high current returns more frequently even if current returns do not predict future ones (Benartzi, 2001; Kaustia and Knüpfer, 2008; Barber et al., 2009; Choi et al., 2009). Similar evidence comes from the literature on heterogeneous expectations (see Hommes, 2011, for an overview of the literature); for example, Chavas (2000) estimates that 47 percent of cattle producers use the current price as proxy for future prices when planning future supply, despite large fluctuations in price over time (widely known as "hog cycle").

The remainder of the paper is structured as follows. In the next section we develop a simple model and derive predictions regarding how current weather may affect advance sales and the subsequent decision of customers whether or not to visit the theatre. In Section 1.3 we describe the data in greater detail. Section 1.4 discusses our main empirical findings. In Section 1.5 we evaluate alternative explanations for our findings as well as their robustness. The last section concludes.

1.2 A Simple Model and Hypotheses

To fix ideas, this section provides a simple model of individual purchase decisions as well as aggregate purchasing behaviour. The model nests rational behaviour as well as projection bias and the "reminder-effect" of weather, where the latter two are models of how the current state – weather on the purchase-date – may affect individual choices and total sales. From the general model, we derive testable predictions to distinguish between rational behaviour and the two potential explanations for weather-dependent individual decisions and sales.

1.2.1 Individual Purchase Decisions and Aggregate Sales

Individual Decisions Survey results indicate that weather is an important determinant of the utility derived from an outdoor movie night.[6] Overall, 81 percent of respondents state that dry weather is "very important" or "important" for having a good night at

[5] Recall that projection bias is a mistake in predicting utility at unknown future states. The beliefs regarding the likelihood of each state are assumed to be correct.

[6] For a description of the survey see Section 1.3.3.

the movie; comfortable temperatures are of importance for 66 percent. In our model, therefore, each customer derives weather-related utility $u(w_\tau)$ when watching a movie on date τ given weather conditions $w_\tau \in \mathbb{R}$.[7] The utility function $u(\cdot)$ is assumed to be increasing, twice differentiable, and concave on the real line. When not visiting the theatre, individuals receive utility $u(\eta)$ from a heterogeneous outside option $\eta \in \mathbb{R}$, which is distributed within the population according to the distribution $G(\cdot)$.

On the purchase-date $t < \tau$, an individual decides whether or not to buy a ticket for the movie-date at costs c (in utility terms). On the purchase-date, the realisation of weather on the movie-date is uncertain. We denote the distribution of w_τ at t by $H(\cdot)$, which is known to potential customers. We assume that $H(\cdot)$ belongs to the location family of distributions with location parameter f_t and is independent of actual weather w_t (we will justify this assumption empirically in Section 1.5.1).[8] The parameter f_t denotes the weather forecast at t for the movie-date τ, which is available to individuals free of charge. The forecast predicts expected weather on the movie-date and contains all relevant information regarding movie-date weather at t: $E[w_\tau | \, f_t] = E[w_\tau | \, f_t, w_t] = f_t$, where E is the expectations operator with respect to $H(\cdot)$.

To incorporate projection bias in our model, we adopt the formulation of "simple projection bias" (Loewenstein et al., 2003) and assume that the current state – in our case current weather – receives weight $\alpha \in [0, 1]$ in an agent's expected utility function. Clearly, the case $\alpha = 0$ represents fully rational behaviour. The case $\alpha > 0$ captures that individuals cannot fully assess the extent to which a change in the state of the world will alter their utility and thus unconsciously anchor their utility on the current state.[9]

Expected utility from purchasing a ticket on the purchase-date for an individual with outside option η is then given by

$$v^B(f_t, w_t, \eta) = (1 - H(\eta)) \Big((1 - \alpha) \, E[u(w_\tau) | \, w_\tau \geq \eta, f_t] + \alpha u(w_t) \Big) + H(\eta) \, u(\eta) - c. \quad (1.1)$$

A customer who owns a ticket will only visit the theatre if movie-date weather exceeds the outside option ($w_\tau \geq \eta$). In this case, captured by the first term of (1.1), she receives weather-related (expected) utility from visiting the theatre, which may have excessive

[7] For simplicity, the model abstracts from potential explanations for ticket orders different from weather such as the popularity of a movie. In the setting we are analysing, these factors are orthogonal to purchase-date weather such that omitting them in this analysis does not alter the empirical implication of the model. We control for popularity of the movie in one of the robustness checks in Section 1.5.3.

[8] In practice, $H(\cdot)$ would depend on the forecast horizon $\tau - t$ and the season of the year as well. Considering these factors does not change the analysis. To ease notation, we therefore omit them here.

[9] As mentioned in the introduction, this interpretation is equivalent to individuals holding beliefs about the distribution of future state, which are unconditionally biased towards the current state.

weight on the current state. If movie-date weather turns out to be unexpectedly bad ($w_\tau < \eta$), she will let her ticket expire and choose the outside option instead. In either case, she has to bear the ticket costs c.

Clearly, an individual with outside option $\bar{\eta}$ will be indifferent between buying and not buying a ticket on the purchase-date iff

$$F^P(f_t, w_t, \bar{\eta}) \equiv v^B(f_t, w_t, \bar{\eta}) - u(\bar{\eta}) = 0 \tag{1.2}$$

A natural candidate for optimal choice behaviour is that all individuals with low outside options $\eta \leq \bar{\eta}$ buy tickets on the purchase-date and all individuals with high outside options $\eta > \bar{\eta}$ do not. Lemma 1.1 below states that optimal choices can indeed be completely described by a unique $\bar{\eta}$ satisfying (1.2). Before stating the lemma, however, we need to assume sufficient conditions for a unique fixed point to exist.

Assumption 1.1.

(i) For all f_t there exists an η satisfying $F^R(f_t, \eta) \equiv (1 - H(\eta)) \Big(E[u(w_\tau)|\, w_\tau \geq \eta, f_t] - u(\eta)\Big) - c = 0$.

(ii) The hazard rate of $H(\cdot)$, $h(w)/(1 - H(w))$, is weakly increasing.

Assumption 1.1 (i) ensures that there is at least one potential customer, who, given the optimal use of information, would be indifferent between buying a ticket on the purchase-date and not buying a ticket at all. This ensures existence of a fixed point of (1.2). Assumption 1.1 (ii) is the monotone hazard rate assumption, which provides a sufficient condition for uniqueness of the fixed point and holds for a variety of frequently used distributions like the normal and uniform distributions. Given this, we can state the following lemma:

Lemma 1.1. *Suppose Assumption 1.1 holds. Then, for each (f_t, w_t) a unique $\bar{\eta}$ satisfying (1.2) exists.*

All proofs are relegated to the appendix. A direct implication of the above lemma is that there is always a positive probability, $G(\bar{\eta}) \in (0, 1)$, that some customer will buy a ticket on the purchase-date.

Aggregate Sales Given the individual propensity to buy a ticket, expected aggregate sales depend on the total number of potential customers. Here, we incorporate the idea in our model that good weather makes the choice option “outdoor movie theatre” more salient and thus enlarges the customer base. One possible interpretation is that customers face cognitive restrictions regarding the number of choice options they can consider at a given time. For this reason, they consider a choice option only if it "comes to mind", which is supposed to be positively related to its attractiveness at the current state.[10]

If the number of potential customers is weather-dependent, ticket sales may be driven by weather even if individual decisions to buy tickets are fully rational. To allow for this explanation in our model, we assume that the number of potential customers $n(w_t)$ is increasing in purchase-date weather w_t. The expected total number of sales on purchase-date t is thus given by $y(f_t, w_t) = n(w_t)\, G(\bar{\eta}(f_t, w_t))$. If customers are fully rational – that is, they have all choice options in mind at all times –, $n(\cdot)$ is independent of w_t.[11]

1.2.2 Hypotheses

Our empirical analysis in Section 1.4 is guided by testable predictions derived from the model. Our first hypothesis deals with the effect of purchase-date weather on sales.

Hypothesis 1.1. *If customers are rational ($\alpha = 0$ and $\partial n(w_t)/\partial w_t = 0$) advance sales are independent of purchase-date weather. Otherwise, sales increase when purchase-date weather is good.*

If we reject the implications of rational behaviour in our data – if variations in purchase-date weather explain variations in advance sales –, our model assumes that this effect can be explained by projection bias or a reminder-effect of current weather. For this to be the case, we expect customers to be unaware of the limitations underlying their choices – otherwise, they could adopt strategies to arrive at optimal choices nevertheless. This conjecture provides a plausibility test for our model, because sales should be affected by purchase-date weather regardless of customer's past experiences or the time horizon between purchase-date and movie-date if customers are indeed unaware of the impact on weather on their purchase decisions.

Furthermore, we derive testable predictions to disentangle whether the current state w_t affects individual decisions via projection bias or whether purchase-date weather solely affects the total number of potential customers. Since we do not observe individuals who abstain from buying a ticket, we answer this question by examining the individual decision to collect paid-for tickets on the movie-date. Our model predicts that individuals buy a

[10] Another possible interpretation for a weather-dependent customer base is that good weather at the purchase-date facilitates the coordination of larger groups.

[11] A third possible explanation for a positive relation between good weather on the purchase-date and sales is that customers expect the theatre to be sold out with higher probability. We discuss this potential explanation theoretically after Hypothesis 1.2 and empirically in Section 1.5.2.

ticket if their outside option is worse than $\bar{\eta}$ and to collect it if movie-date weather is sufficiently nice ($w_\tau > \eta$). The probability that a customer collects her ticket is therefore given by

$$\Pr(\text{collect} \mid \text{buy}) = 1 - \frac{\Pr(w_\tau < \eta < \bar{\eta})}{\Pr(\eta < \bar{\eta})} = \begin{cases} G(w_\tau)/G(\bar{\eta}) & \text{if } w_\tau < \bar{\eta} \\ 1 & \text{else.} \end{cases} \tag{1.3}$$

If individual purchase decisions are affected by current weather, good weather increases the expected utility of buying tickets in advance and thus leads to a higher $\bar{\eta}$. Since the realisation of movie-date weather w_τ is independent of purchase-date weather w_t, the likelihood that a customer prefers her outside option therefore increases if purchase-date weather was nice. In contrast, if current weather has no effect on individual decisions (but only on the aggregate number of customers), the likelihood of ticket collection is expected to be independent of purchase-date weather. The following hypothesis summarizes this argument.

Hypothesis 1.2. *If customers are rational ($\alpha = 0$ and $\partial n(w_t)/\partial w_t = 0$) or if current weather increases the pool of potential customers ($\partial n(w_t)/\partial w_t > 0$), the probability that tickets are collected is independent of purchase-date weather. Otherwise – if individual decisions are affected by projection bias ($\alpha > 0$) and if movie-date weather is worse than expected – the probability that tickets are collected decreases when purchase-date weather is good.*

Before we continue, it is important to point out a few assumptions upon which our model and hypotheses rests. First, as noted above, we assume that purchase-date weather has no information value for movie-date weather. Otherwise, our results could be explained by customers taking current weather as informative signal. We show in Section 1.5.1 that purchase-date weather is indeed not informative. Nevertheless, individuals could perceive current weather to be informative for the future. As discussed in the introduction, we cannot rule out this explanation if individuals perceive purchase-date weather to be informative regardless of the time span between purchase-date and show. However, it is natural to assume that the perceived information content of current weather is declining in the time horizon one is trying to predict. Then, we would expect that the effect of purchase-date weather on sales becomes weaker with increasingly long horizons. In Section 1.5.2 we will see that this is not the case.

Finally, to keep the model simple, we have abstracted from the fact that potential customers essentially face a dynamic problem when they decide on which date they would like to buy their tickets. Clearly, the timing of buying tickets can be affected by purchase-date weather, for example if the latter affects the perceived probability that the theatre may sell out. We discuss this potential alternative explanation in more detail in Section 1.5.2.

1.3 Data

Our data comes from four different sources. An outdoor movie theatre located in Munich, Germany, provided us the record of their online advance tickets sales platform. The Meteorological Institute of the University of Munich shared their detailed data on the weather conditions in Munich with us; the local weather forecast was collected from the archives of the newspaper “Süddeutsche Zeitung”, a high quality newspaper located in Munich. Finally, we conducted a survey among visitors of the theatre at 13 different nights of the 2011 season.

1.3.1 Weather and Forecast

We collect data on weather and weather forecasts for the months June to August of the years 2004 to 2011, which are the times at which the theatre screens and for which we have sales data (for details, see below).

The Meteorological Institute of the University of Munich[12] provides us hourly measures for precipitation (measured in 1/100 mm), temperature (measured in degrees Celsius) as well as the average sunshine duration (in percent) between 8 am and 7 pm.[13] Most statistical inferences uses daily averages (24 hours) of these three weather variables.

We hand-collect the weather forecast from the archives of the daily newspaper “Süddeutsche Zeitung”, which is published every day except Sundays and public holidays.[14] It provides a regional forecast for each day, one to four days out, for the South of Bavaria including Munich. The forecast comprises forecasted maximum and minimum temperature (in degrees Celsius) and one of the following weather symbols: sunny, partly sunny, shower, rain, and scattered thunderstorms.[15]

The weather in Munich is highly variable, especially during the summer months, during which there are daily shows at the movie theatre. This is mostly due to the proximity

[12] The distance of the weather station to the movie theatre is 5.3 km (3.3 miles).

[13] The latter restriction ensures that the changing times of dusk and dawn do not confound our measure of sunshine duration.

[14] Weather forecasts take up a lot of memory capacity, which is why they are not stored by any German weather firm.

[15] There are in total 12 observations of the symbol overcast, which we group with “shower” to simplify the exposition of results. Undoing this grouping does not lead to any significant changes throughout.

Table 1.1: Summary Statistics: Weather and Forecast

	Weather		
	All day	Evening	SD within day
Avg. Sunshine Duration	54.32	47.86	18.73
	(34.99)	(38.06)	(13.96)
Avg. Temperature	19.01	19.15	3.01
	(3.60)	(3.99)	(1.28)
Avg. Rainfall	11.23	20.27	28.38
	(24.81)	(63.16)	(62.83)
	Forecast		
	Minimum	Maximum	
Forecasted Temperature	12.66	23.56	
	(2.75)	(4.02)	

Notes: We report the means of variables; their standard deviations are in parentheses. Sunshine duration is mesured in percent per hour, temperature is measured in degrees Celsius, and rainfall is reported in 1/100 mm per hour. In the column "SD within day" we report the average of the variable's standard deviations over the course of each single day.

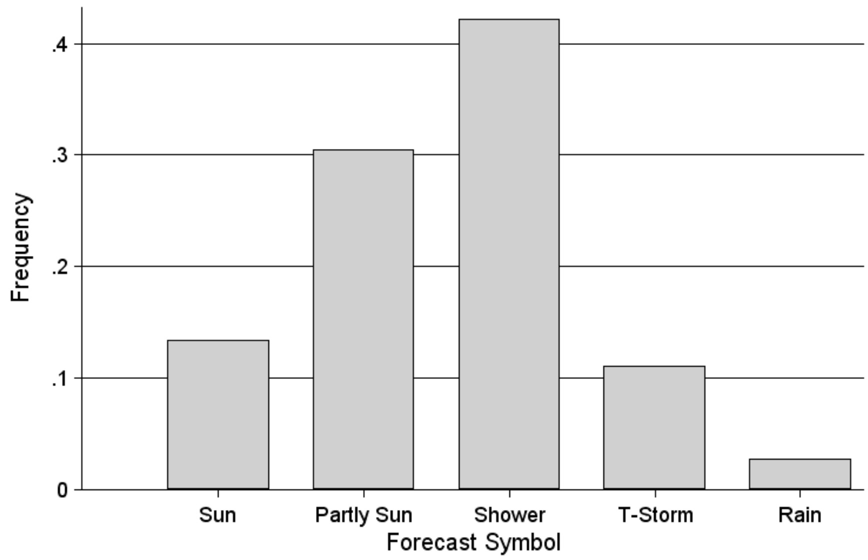
of the Alps, which leads to frequent and often unexpected rainstorms. These tend to occur especially in the evening hours. For this reason, there is high monthly precipitation in the summer months, when total precipitation is on average 123 mm per month (for comparison: London 51 mm, New York City 92 mm, and Berlin 61 mm). Long periods of stable good weather are the exception; rather, there are frequent shifts in weather patterns every few days as reflected by the mean number of 12.4 rain days per month (days with at least 1 mm of rain) during the summer months (for comparison: London 10.5 days, New York City 8 days, and Berlin 8.7 days).[16]

The weather thereby varies within as well as across days. This can be seen in Table 1.1, where the summary statistics of average daily weather are depicted in the first two columns. Standard deviations of sunshine duration as well as rainfall are high compared to their respective means; the coefficient of variation for sunshine duration is 1.55 and 0.45 for rainfall. Furthermore, it is noteworthy that rainfall in the evening hours is considerably higher than during the day reflecting the higher likelihood of rainstorms at these times. The third column of Table 1.1 provides information for the variation of weather within days by depicting the mean of within-day standard deviations of the respective weather variable. Note that both sunshine duration and precipitation exhibit high within-day variation. The within-day variation for temperature is not very informative, as there is a cyclical pattern of temperature within each day. (Keep in mind that sunshine duration

[16]Sources of long term monthly averages: World Meteorological Organization http://worldweather.wmo.int/ [4 October 2012].

Figure 1.1: Distribution of Forecasted Weather (Symbols)

This figure plots the distribution of forecast symbols pooling over forecast horizons (one to four days in advance).



is only measured between 8 am and 7 pm such that darkness does not contribute to the within-day variation of sunshine duration.)

Regarding the forecast, note first that average forecasted temperatures are in a similar range as average temperatures (Table 1.1), which is what we expect. The distribution of weather symbols for all forecast horizons – as shown in Figure 1.1 – again reflects the high variations in local weather across days.[17] Note furthermore that the forecast frequently predicts scattered thunderstorms and showers, which indicates rather unstable weather conditions within days as well.

1.3.2 Ticket Sales

The data on advance ticket sales were provided by "Kino, Mond und Sterne" [Movies, Moon, and Stars], one of four outdoor movie theatres in Munich. The theatre usually screens daily during the months of June, July, and August, and shows the movie regardless of weather conditions. The latter fact is important for our study, since it implies that tickets bought in advance are non-refundable. A consumer, who buys a ticket for this theatre in advance, thus bears the full weather risk.[18] Customers are expected to be aware of this risk, as it is explicitly mentioned prominently on the ticketing website.

The theatre has a total of 1,300 seats available, tickets for which are sold at the box office

[17]The distributions of symbols by forecast horizon do not differ substantially.

[18]None of the seats are covered. See Appendix A.3 for a picture of the theatre.

and various advance ticket sales locations. The majority of advance sales are sold online where tickets for a particular show are available until 6 pm on the day of the screening. Our goal is to explain these advance sales such that our main data set comprises of all online ticket orders for the theatre between 2004 and 2011. This amounts to a total of 20,999 orders.[19] For each order, the system records the number of tickets bought, the exact date of the transaction and a unique alphanumeric customer ID, which allows us to track repeat customers.

Additionally to the data on online sales for the years 2004 to 2011, we have data on the total number of visitors of the theatre – including box office sales – for the years 2009 to 2011. This allows us to assess the importance of advance online sales, which amount to 24 percent of the total number of tickets during this period. More than half (almost 60 percent) of online tickets are sold on the day of the show. Our main analysis focuses on sales between one and four days before the show, on which the weather forecast for the movie-date is available. Within this period, 30 percent of online tickets are sold, with percentages declining between one and four days out. The remaining 10 percent of online tickets are sold five days or earlier before a show.

Our main variable of interest is aggregate ticket sales on a daily base. More precisely, one observation is the sum of ticket orders on a single day for a specific show. If no tickets are sold on a day at most 23 days before the show, we add an observation with aggregate orders of zero. This results in at least 24 observations for every single movie shown, one for each day between 0 and 23 days out. We construct additional aggregates of ticket sales for robustness checks. For example we count orders of repeat customers (identified by their unique customer ID), who have bought tickets more than once since 2004. Another noteworthy variable are ticket orders by repeat customers who had previously bought tickets for a show during which it was raining.

For the years 2009 – 2011, we additionally know for each order whether tickets were in fact collected at the evening of the show. Of the total of 4,102 orders, the vast majority (88 percent) of tickets was collected on the day of show.

The summary statistics for ticket sales are presented in Table 1.2, organised according to how early in advance tickets were sold. The average number of ticket orders decreases from 7 one day out to 1 four days out, representing the declining pattern of orders. The number of tickets sold per order remains stable at about 2.6, independent of the time horizon. About half of the ticket orders are placed by repeat customers, who have bought tickets online more than once.

[19] Ticket prices have been stable at about 7.40 Dollar (5.70 Euro) each during the entire period.

Table 1.2: Summary Statistics: Orders

	Day of show	1 day out	2 days out	3 days out	4 days out
Avg. Orders	24.74	7.18	2.78	1.36	0.89
	(33.97)	(10.82)	(4.30)	(2.24)	(1.37)
Tickets per Order	2.46	2.55	2.55	2.63	2.58
	(0.69)	(0.92)	(0.88)	(1.24)	(1.24)

Notes: We report the means of daily ticket orders bought on the day of the show as well as one to four days in advance. For the same five days we additionally provide average numbers of tickets per order. Standard deviations in parentheses.

1.3.3 Survey

During the 2011 season, we conducted a survey among visitors of the cinema. As many visitors spend some time in the theatre before the movie starts, the willingness to participate in the survey was high. Overall we received 443 questionnaires for 13 different days with considerable variance in weather conditions (and accordingly varying number of questionnaires obtained per day). This amounts to more than 10 percent of the audience on these days on average. Of all surveyed customers, 25 percent bought their ticket online (compared to 24 percent of all customers in the years 2009 – 2011) and 7 percent purchased it one to four days in advance (compared to 8 percent of all customers between 2009 and 2011). Throughout, we use the survey to provide supporting evidence for our arguments. That being said, none of our main results depends on data from the survey.

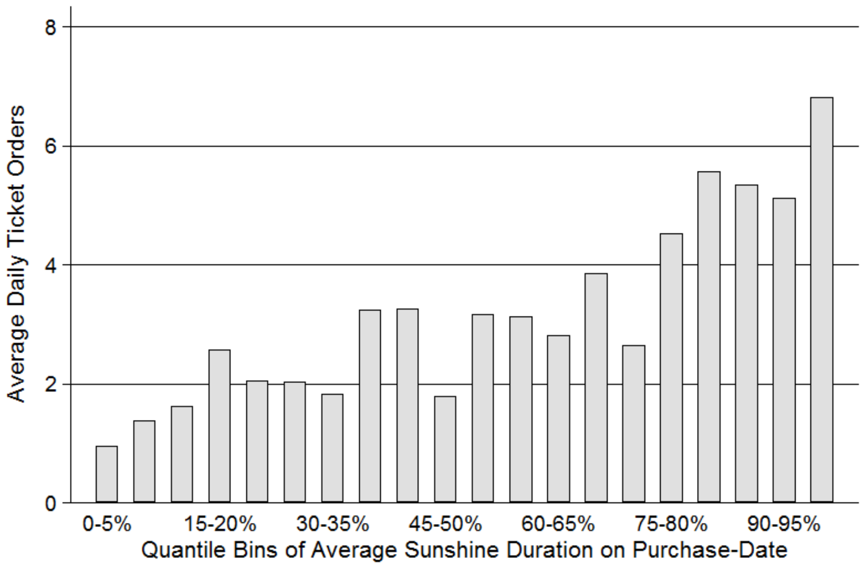
1.4 Empirical Analysis

In this section we test the hypotheses derived in Section 1.2. We first show that weather on the purchase-date explains variation in ticket orders for various model specifications, rejecting rational behaviour from Hypothesis 1.1. Furthermore, we show that good weather on the purchase-date decreases the likelihood that the purchased tickets are collected on the movie-date, providing evidence for projection bias (Hypothesis 1.2).

1.4.1 Purchase-Date Weather and Ticket Orders

Figure 1.2 illustrates the effect of weather on ticket orders by comparing the number of orders across different weather conditions on the purchase-date. For the sample of ticket orders one to four days ahead of the movie date – which we use in the empirical analysis below – we group ticket orders into bins based on five percent quantiles of purchase-date

Figure 1.2: Purchase-Date Weather and Ticket Orders
This figure plots the average number of daily ticket orders (between one and four days in advance) for bins based on five percent quantiles of purchase-date sunshine duration. Bins are sorted from dates with shortest sunshine duration (to the left of the horizontal axis) to days with longest sunshine duration (to the right of the horizontal axis).



sunshine duration and plot, for each of these bins, the average number of ticket orders per day. Consistent with projection bias, Figure 1.2 shows that the average number of daily ticket orders strongly rises parallel to an increase in sunshine duration from the left to the right of the horizontal axis.

An obvious concern with the graphical analysis above is that other factors which explain ticket sales – like the weather forecast – may possibly be correlated with purchase-date weather. To address this concern, we estimate the effect of purchase-date weather on the number of daily ticket orders in a number of regressions. In all of these regressions, we include average sunshine duration as well as average precipitation on the purchase-date t as explanatory variables (collected in the weather vector $\mathbf{W_t}$).[20] In addition, we control for the weather forecast at t for the movie-date τ by adding the forecasted maximum and minimum temperatures, as well as separate dummy variables for each forecast symbol as independent variables; these variables are collected in the forecast vector $\mathbf{F_{\tau t}}$. Because the forecast is only available for a horizon Δ of up to four days, we limit the sample to ticket orders between one and four days ahead of the show.

For the first empirical model we organise the data in a panel structure with day of the

[20]We omit average temperature from our analysis, since it is highly correlated with sunshine duration ($\rho = 0.6$), which makes the analysis of the respective coefficients difficult. We chose to keep sunshine duration for its greater salience compared to temperature. However, our results are not qualitatively affected by this choice. For further details see the discussion in Section 1.5.3.

show τ as unit of observation and advance sales one to four days out being observations over time. Within this structure the model is

$$y_{\tau t} = \mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{W}} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{D}'_{\tau\mathbf{t}}\beta_{\mathbf{D}} + \upsilon_{\tau t}, \tag{1.4}$$

where $\mathbf{D}_{\tau\mathbf{t}}$ includes dummy variables for each time difference between purchases and show. We assume that the error term $\upsilon_{\tau t}$ is iid between different shows τ but may be arbitrarily correlated between advance sales for the same show. To control for unobserved heterogeneity possibly correlated with our regressors, we estimate (1.4) as a fixed effects model.

For the second econometric model we organise our data as cross sections separately for each purchase-date being $\Delta \in \{1, 2, 3, 4\}$ days ahead of the movie-date. This gives us less power due to limiting observations, but allows us to exploit cross sectional variation and to include a set of controls $\mathbf{X}_{\tau\mathbf{t}}$, which are for most observations time invariant between t and τ. Specifically, we control for the day of the week of the show τ, average sunshine duration and precipitation of the past two weeks before t, as well as dummy variables for year and month. For each Δ, we estimate the following model:

$$y_{\tau t} = \mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{W}} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{X}'_{\tau\mathbf{t}}\beta_{\mathbf{X}} + \varepsilon_{\tau t}. \tag{1.5}$$

Since $\Delta = (\tau - t)$ is fixed, there is a single observation for each movie night τ. Imposing the identifying assumption from above – that errors $\varepsilon_{\tau t}$ are uncorrelated across movie-dates τ – we can estimate (1.5) by OLS.

Table 1.3 displays the estimation results for the two models. Similar to the graphical analysis in Figure 1.2, the results provide strong support for Hypothesis 1.1: the effect of sunshine duration on aggregated ticket sales is positive and significant throughout. In the fixed effects model (1.4) and the cross sections (1.5) for one and three days out, average rainfall has furthermore a negative effect on ticket sales, which is significant at least at the ten percent level. Moreover, as predicted by the theoretical model, we find significant effects of the weather forecast on sales at least for one to three days out. This is especially true for forecasted temperature. Forecast symbols seem to have an effect one and two days out, only. In the fixed effect model (1.4) we find no statistically significant effect of these symbols at all, which may mostly be explained by their limited within variance.

In order to interpret the estimated parameters of the variables of interest and to compare their impact across different advance sales horizons Δ, we calculate the statistic $m(x) = \beta(x)\, s(x)/\bar{y}_{\Delta}$ for the estimated models. The nominator of $m(x)$ is the product of the coefficient $\beta(x)$ of an independent variable x and its standard deviation $s(x)$, which gives

Table 1.3: Effect of Purchase-Date Weather on Ticket Orders

	Daily Ticket Orders				
	Fixed Effects	1 day out	2 days out	3 days out	4 days out
Avg. Sun	0.023***	0.027**	0.016***	0.0061*	0.0046**
	(0.0054)	(0.014)	(0.0054)	(0.0032)	(0.0021)
Avg. Rain	−0.014**	−0.028**	0.013	−0.0059*	0.0020
	(0.0063)	(0.014)	(0.012)	(0.0036)	(0.0033)
Forecasted Maxtemp.	0.19**	0.68***	0.26***	0.15***	0.053*
	(0.072)	(0.17)	(0.084)	(0.041)	(0.027)
Forecasted Mintemp.	−0.014	0.61**	0.12	−0.0063	0.019
	(0.090)	(0.24)	(0.12)	(0.056)	(0.030)
Symbol Partly Sunny	0.30	−4.05*	−2.31**	0.32	−0.55**
	(0.79)	(2.19)	(0.94)	(0.40)	(0.24)
Symbol Shower	−0.21	−6.92***	−3.30***	−0.17	−0.46*
	(0.74)	(2.23)	(0.90)	(0.37)	(0.24)
Symbol Rain	−1.13	−8.09***	−3.19***	0.49	−0.61
	(1.02)	(2.62)	(1.14)	(0.59)	(0.50)
Symbol T-Storm	−0.53	−10.7***	−1.95*	0.34	−0.12
	(0.81)	(2.32)	(1.04)	(0.47)	(0.36)
2 Days Out	−4.44***				
	(0.38)				
3 Days Out	−5.94***				
	(0.47)				
4 Days Out	−6.43***				
	(0.49)				
Time-invariant Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1635	413	411	406	405
Adjusted R^2	0.282	0.350	0.306	0.205	0.181

Notes: We report the coefficients and robust standard errors of OLS regressions of total daily ticket orders on purchase-date weather, forecast, and control variables. In the first column, the sample consists of all daily ticket orders between one and four days before the show (show-date and horizon fixed effects included). In the remaining columns, the sample is split according to the number of days out tickets are purchased. "Avg. Sun" is the average sunshine duration in percent on the purchase-date between 8 am and 7 pm; "Avg. Rain" denotes average rainfall on the purchase-date in 1/100 mm. Forecasted temperatures are from the forecast at the purchase-date for the movie-date and measured in degrees Celsius. The variable "Symbol Partly Sunny" takes the value 1 if the forecast is partly sunny for the movie-date on the purchase-date and 0 otherwise. Other symbol-variables are defined accordingly; the baseline forecast symbol is sunny.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

us the impact of one standard deviation change of x on the number of advance sales. To compare this effect across different advance sales horizons Δ, we normalize it by the mean of the respective number of advance sales $\bar{y}_{\Delta}$. Thus, the statistic $m(x)$ denotes the impact of a one standard deviation change of x on sales as percentage of mean sales for a given empirical specification.

In the estimated model, one standard deviation change of actual sunshine duration leads to a change in sales between 10 and 25 percent of the mean. In comparison, one standard deviation of the forecasted temperature has an effect on sales between 10 and 40 percent of the mean – the effects of these two determinants of sales are thus of comparable size. This leads to our first result.

Result 1.1. *Purchase-date weather has a statistically and economically significant effect on aggregate ticket orders.*

We further investigate the conjecture from Section 1.2.2 that the effect of weather on sales is independent of customers' past experiences to address the concern that the results are driven by inexperienced customers.[21] To this end, we estimate the fixed effects model (1.4) replacing total advance sales on a given day by sales to three different subsets of repeat customers as the dependent variable.

The results of this exercise are reported in Table A.1 in Appendix A.2.[22] Average sunshine duration has a positive and highly significant effect on sales to the sub-population of repeat customers, who had bought tickets at least previously once during the entire period between 2004 and 2011 or at least once during the same season. Ticket orders by customers, who had previously bought tickets for a show during which it was raining, can also be explained by variations in sunshine duration on the purchase-date. The economic significance as measured by statistic $m(\cdot)$ is in approximately the same range as the estimates for the complete sets of customers for both forecast and purchase-date weather. These results suggest that decisions of experienced customers are influenced by current weather, even if weather at previous visits turned out to be bad.

1.4.2 Purchase-Date Weather and Ticket Collection

So far, our analysis has focused on explaining aggregate purchase behaviour. Using aggregate data, we cannot distinguish whether the current state – purchase-date weather

[21] We investigate the additional conjecture that the effect of weather is independent of the time horizon between purchase-date and movie-date in Section 1.5.2

[22] Although we only report the results from the fixed effect model (1.4), similar results are obtained from model (1.5). The results can be obtained from the authors on request.

– affects the number of potential customers (through reminding them of their choice options) or whether it affects individual decision making directly (through projection bias). We assess to what extent purchase-date weather alters individual behaviour by exploiting information on the individual decision whether or not to collect paid-for tickets on the movie-date.

According to the model in Section 1.2, the likelihood that tickets are collected decreases with good weather on the purchase-date if individual decisions are directly affected. In contrast, if good weather at purchase merely increases the number of potential customers, the decision to actually visit the theatre is expected to be independent of purchase-date weather.

Let the decision of a customer i, who has purchased a ticket on date t for a movie at τ, to collect her ticket be denoted by $\psi_{it\tau} = 1$ and the decision to let the ticket expire by $\psi_{it\tau} = 0$. The likelihood that a customer collects her ticket on the movie-date is estimated using the probit model

$$Pr(\psi_{it\tau} = 1) = \mathbf{\Phi}(\mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{Wt}} + \mathbf{W}'_{\tau}\beta_{\mathbf{W}\tau} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{X}'_{\tau\mathbf{t}}\beta_{\mathbf{X}}). \tag{1.6}$$

Since the model predicts that individual collection decisions depend on movie-date weather, we include movie-date weather $\mathbf{W}_{\tau}$ on the right hand side of (1.6) additional to purchase-date weather $\mathbf{W_t}$, forecast $\mathbf{F}_{\tau\mathbf{t}}$, and controls $\mathbf{X}_{\tau\mathbf{t}}$ as defined in Section 1.4.1. Since sunshine duration ceases to be a salient indicator for actual weather at night, we add a dummy variable indicating whether tickets were purchased later than 8 pm to the vector $\mathbf{X}_{\tau\mathbf{t}}$. In order to assess the robustness of the estimates, we re-estimate (1.6) without controls $\mathbf{X}_{\tau\mathbf{t}}$.

The first two columns of Table 1.4 report the estimated coefficients from model (1.6) for all customers who had purchased tickets one to four days in advance in the years 2009 to 2011. Extended sunshine duration on the purchase-date tends to reduce the likelihood that tickets sold in advance are actually collected at the box office. However, the estimated coefficients are not significantly different from zero at any common level. We conjecture that this is due to limited variance of the dependent variable: 93 percent of all customers collect their ticket. In fact, the model predicts that the probability for advance tickets being collected equals one if the realised weather at the movie turns out to be at least as good as the outside option of the marginal customer. In other words, our model predicts that customers should decide to let their tickets expire only in cases in which movie-date weather is a negative surprise. Including all customers in our analysis should therefore bias the coefficient for purchase-date sunshine duration upwards.

To test this conjecture, we estimate (1.6) for a sample of movie-dates at which realised

Table 1.4: Effect of Purchase-Date Weather on Ticket Collection

	Tickets Collected			
	(1) Full Sample	(2) Full Sample	(3) Restricted Sample	(4) Restricted Sample
Avg. Sun	−0.0020	−0.0019	−0.0049**	−0.0049*
	(0.0016)	(0.0017)	(0.0025)	(0.0026)
Avg. Rain	0.00080	0.00062	0.0024	0.0000020
	(0.0023)	(0.0025)	(0.0034)	(0.0037)
2 Days Out	−0.28***	−0.30***	−0.28**	−0.39***
	(0.10)	(0.11)	(0.14)	(0.15)
3 Days Out	−0.084	−0.13	0.057	0.0036
	(0.14)	(0.15)	(0.18)	(0.20)
4 Days Out	−0.52***	−0.61***	−0.40*	−0.46*
	(0.16)	(0.16)	(0.23)	(0.24)
Sun before Film	0.0067***	0.0065***	0.012**	0.016***
	(0.0015)	(0.0017)	(0.0047)	(0.0046)
Rain Film	−0.0011**	−0.00099*	−0.00098*	−0.0018**
	(0.00050)	(0.00056)	(0.00059)	(0.00076)
Temp. Film	0.15***	0.15***	0.17***	0.16***
	(0.020)	(0.021)	(0.025)	(0.028)
Forecasted Maxtemp.	−0.082***	−0.086***	−0.047*	−0.057*
	(0.023)	(0.026)	(0.028)	(0.032)
Forecasted Mintemp.	−0.024	−0.018	−0.065*	−0.061
	(0.028)	(0.029)	(0.036)	(0.039)
Symbol Partly Sunny	0.10	0.094	0.25	0.28
	(0.13)	(0.14)	(0.20)	(0.24)
Symbol Shower	0.0059	−0.012	0.10	0.12
	(0.14)	(0.15)	(0.23)	(0.25)
Symbol T-Storm	0.21	0.22	0.34	0.58**
	(0.17)	(0.18)	(0.24)	(0.27)
No. of Tickets	0.097**	0.10**	0.046	0.051
	(0.043)	(0.044)	(0.043)	(0.044)
Evening	0.19*	0.20**	0.11	0.13
	(0.097)	(0.097)	(0.14)	(0.14)
Controls	*No*	*Yes*	*No*	*Yes*
Observations	2620	2620	874	874
Pseudo R^2	0.220	0.227	0.200	0.218

Notes: We report the coefficients and robust standard errors of Probit regressions explaining individual decisions to collect advance tickets. The first two columns refer to the full sample of all advance sales for the seasons 2009 – 2011. Column (3) and (4) provide estimates for a restricted sample considering only shows during which the weather was worse than expected (see the text for details). Variables indicating purchase-date weather ("Avg. Sun" and "Avg. Rain") and variables of the purchase-date weather forecast ("Forecasted Max(Min)temp.", "Symbol ...") are defined as in Table 1.3. The baseline forecast is sunny. We control for movie-date weather by "Sun before Film" (average sunshine duration between 5 pm and 7 pm in percent), "Rain Film" (rainfall in 1/100 mm between 7 pm and 11 pm), and "Temp. Film" (temperature on the movie-date between 7 pm and 11 pm), as well as for whether the ticket was bought at darkness on the purchase-date (Evening = 1), the number of tickets ordered ("No. of Tickets"), and the number of days tickets were bought in advance ("x Days Out").
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

weather was a negative surprise. For all movie-dates in the restricted sample, expected sunshine duration $\hat{S}_t$ as predicted by model (1.7) in Section 1.5.1 is greater than realised sunshine duration shortly before the movie starts. As predicted by the theoretical model, almost all customers (97 percent) dropped from the sample chose to pick up their tickets; in contrast, 15 percent of the remaining customers let their ticket expire.

The results from re-estimating (1.6) for this sample are depicted in columns (3) and (4) of Table 1.4. It becomes apparent that the magnitude of the (negative) effect of purchase-date sunshine duration on the likelihood of tickets to be collected doubles at the mean of the sample. This confirms our conjecture that the coefficients in columns one and two are upward biased.[23] Furthermore, the coefficients become statistically significant at the five and ten percent level, respectively. This leads to our second result.

Result 1.2. *Customers are less likely to collect their tickets on the movie-date if they experienced good weather on the purchase-date, providing evidence for projection bias (Hypothesis 1.2).*

1.5 Alternative Explanations and Robustness

In the previous section we have shown that projection bias can account for both, weather-dependent sales and decisions to collect purchased tickets. However, there may exist alternative explanations that could explain these findings as well. In this section, we discuss plausible alternative explanations and the robustness of our results to different empirical specifications.

1.5.1 Is Current Weather Informative for Future Weather?

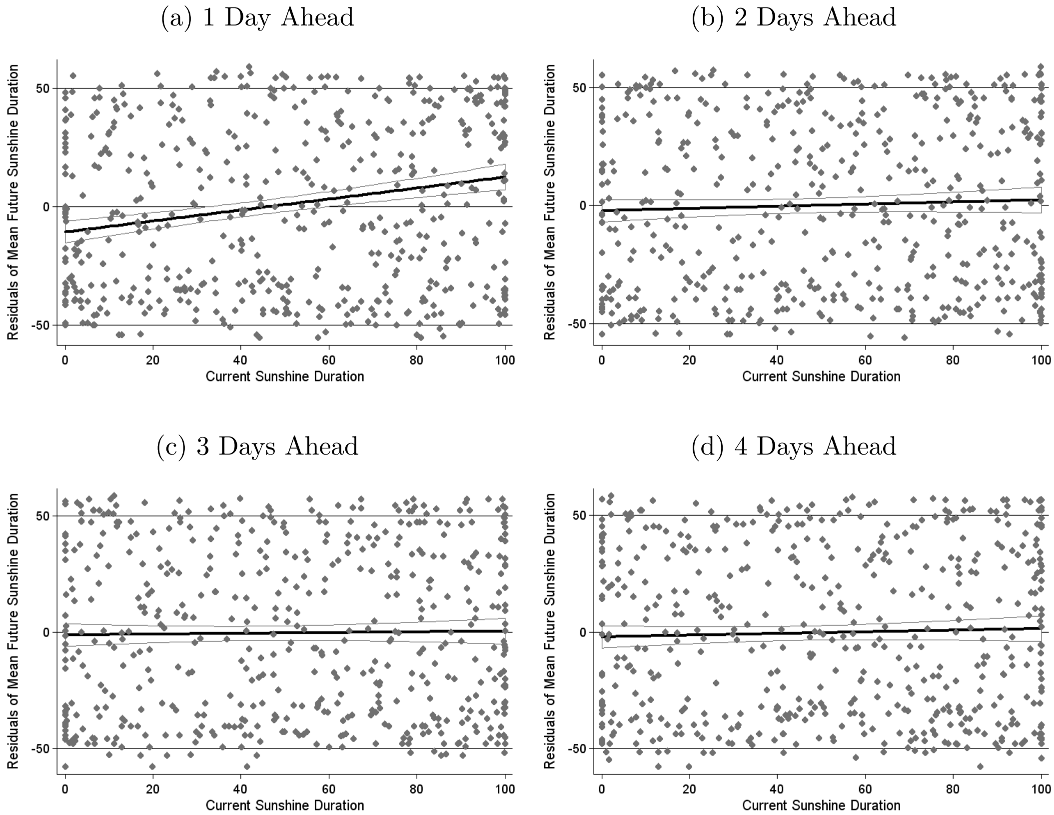
An immediate concern of our analysis so far is that individuals use current weather to update their beliefs about future weather conditions. There are two reasons why this could be optimal. First, current weather may be informative by itself such that looking up the weather forecast is unnecessary. Second, current weather may enhance the prediction of future weather, even given the weather forecast. This may be the case, for example, if the forecast cannot take regional factors into account sufficiently well.

We argue that the information content of current weather for future weather is, in general, limited if not nil due to large day to day fluctuations of local weather in Munich. In Figure 1.3, we plot average sunshine duration one to four ahead (purged for seasonal

[23] The estimates from the restricted sample are potentially still upward biased if customers are subject to the sunk cost fallacy (Arkes and Blumer, 1985).

Figure 1.3: Predictive Power of Current for Future Sunshine Duration

This figure provides a scatterplot of current sunshine duration against residuals of a regression of future sunshine duration (1 to 4 days ahead, respectively) on month and year dummy variables. The black solid line depicts the estimates of a regression of residuals on current weather, the grey lines depict the 95 percent confidence interval.

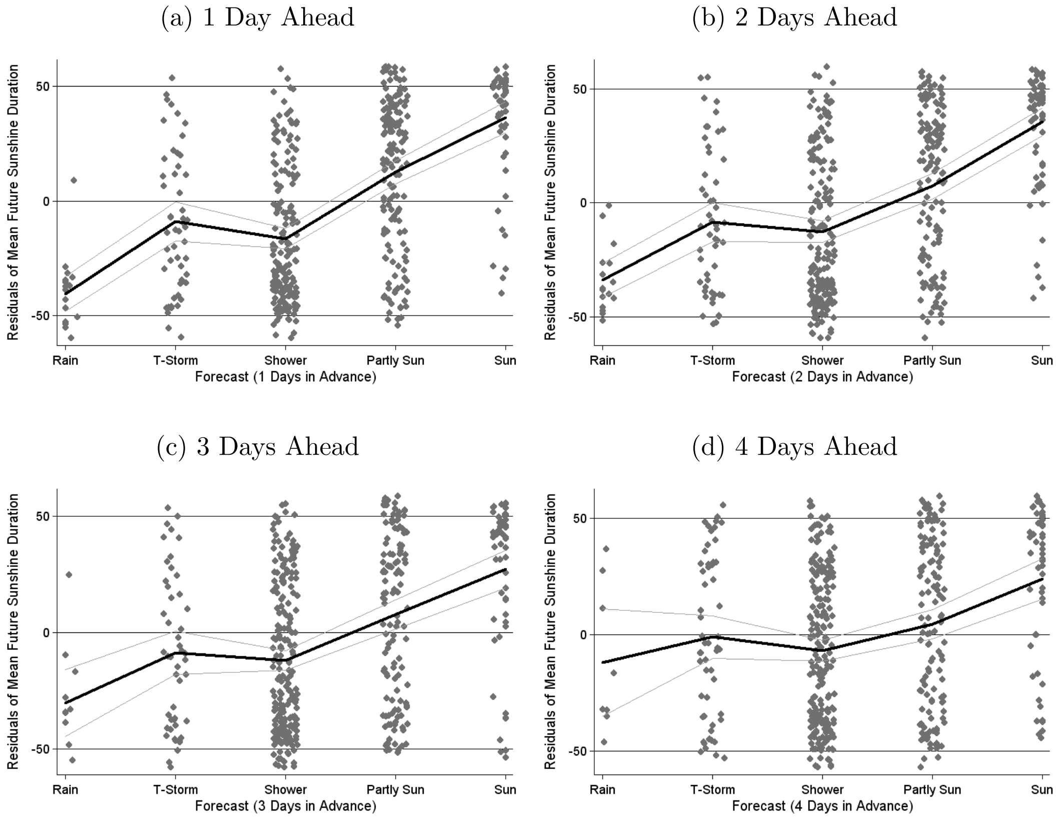


effects by year and month dummies) against current sunshine duration. It turns out that tomorrow's sunshine hours are at best slightly positively related to today's sunshine duration. Furthermore, today's weather has no explanatory power for weather two or more days out.

In contrast, the weather forecast is able to explain future sunshine duration well. Figure 1.4 again plots average sunshine duration purged for seasonal effects as above but this time against the forecast as given by forecast symbols. Evidently, there is a clear positive relationship between symbols indicating good future weather and realised sunshine duration. Especially, the symbols "rain", "partly sunny", and "sunny" seem to predict the weather quite well, even as far as four days into the future.

In order to reassure that the predictive power of current weather – even when not controlling for the forecast – is low, we complement the graphical analysis above with empirical estimates. In particular, we forecast evening sunshine duration S_h (in percent) at some date $h >$ t with the following model

Figure 1.4: Predictive Power of the Weather Forecast for Future Sunshine Duration
This figure provides a scatterplot of current forecast symbols against residuals of a regression of future sunshine duration (1 to 4 days ahead, respectively) on month and year dummy variables. The black solid line connects the means of future sunshine conditional on the forecast, the grey lines connect the 95 percent confidence intervals of the conditional means.



$$S_h = \mathbf{W_t'}\gamma_{\mathbf{W}} + \mathbf{F_{ht}'}\gamma_{\mathbf{F}} + \mathbf{V_t'}\gamma_{\mathbf{V}} + \xi_{ht}, \tag{1.7}$$

where controls $\mathbf{V_t}$ include average sunshine duration and precipitation of the past two weeks before t as well as year and month dummies. Current weather $\mathbf{W_t}$ and forecast indicators $\mathbf{F_{ht}}$ are defined as above. We estimate model (1.7) with and without including the forecast $\mathbf{F_{ht}}$; the results including the forecast are displayed in Table A.2, the results without forecast in Table A.3 in Appendix A.2. Confirming the graphical results, current weather does not help to assess future weather except for one day ahead where the coefficients of current sunshine duration are statistically significant but small (a one percent increase in sunshine duration today leads to an increase in sunshine duration tomorrow of at most 0.21 percentage points). In contrast, the predictive power of the forecast is sizable, since adding it to the model leads to a roughly threefold increase in variance explained. To further appreciate the predictive power of weather symbols, note that evening sunshine duration has a standard deviation of 38. If the forecast symbol for four days in advance is “shower” instead of “sunny”, evening sunshine duration decreases by 70 percent of one standard deviation.[24]

Given this, the question arises whether customers appreciate the predictive power of the forecast. Our survey results indicate that this is indeed the case as customers report to consult the weather forecast frequently and appreciate its reliability. From all respondents, 84 percent consult the weather forecast at least every other day or when they are planning weather related activities. Regarding forecast reliability, 85 (86) percent state that the forecast for tomorrow (two days ahead) will be correct in at least 80 (60) percent of cases.

Overall, given the above results, it seems unlikely that customers knowingly base their predictions of future weather-dependent utility on actual weather – especially since the vast majority of customers are locals, who should be expected to know the regional weather conditions well.

1.5.2 Probability of Ticket Availability

Another concern is that the theatre has a capacity constraint of 1,300 seats such that higher ticket sales at any given point in time lead to a higher risk that the movie may sell out. Thus, if customers believe that the likelihood that tickets will be available on the movie-date decreases with good weather on the purchase-date, they have a higher incentive to buy on the purchase-date. Such a "precautionary" motive for buying tickets early at times of good weather would shift purchase decisions to earlier dates with good weather, which could be a potential explanation for our results.

In fact, the movie theatre in question has been sold out in 13 percent of evenings over the entire time span of our analysis, but has, so far, never been sold out in advance. In general, customers seem to understand that they are always able to buy tickets online: 88 percent of customers state that it is "unlikely" or "very unlikely" that all tickets for tomorrow's screening will be sold out in advance.[25]

Some customers may nevertheless perceive the probability of ticket availability to depend on current weather few days before the show. However, deferring the purchase decision to a later date should perceived to be riskless for particularly early purchase-dates, for instance five days in advance and earlier. Thus, if customers' concerns that the theatre may sell out were the sole explanation for the effect of current weather on sales, particularly early ticket orders should be unaffected by purchase-date weather. In contrast, if our results

[24] Here, we assess the predictive power of current weather and the weather forecast for the main weather indicator of interest – sunshine duration – only. Repeating the exercise for precipitation and temperature gives similar results.

[25] The likelihood that the theatre sells out, however, is empirically unrelated to weather (and the forecast) prior to the movie-date.

Table 1.5: Effect of Purchase-Date Weather on Early Ticket Orders

	Daily Ticket Orders			
	5 - 11 Days Out	9 - 15 Days Out	13 - 19 Days Out	17 - 23 Days Out
Avg. Sun	0.0012***	0.00091***	0.00041	0.00058**
	(0.00043)	(0.00034)	(0.00026)	(0.00023)
Avg. Rain	−0.00046	0.00036	−0.00014	−0.00022
	(0.00043)	(0.00034)	(0.00029)	(0.00021)
Horizon Indicators	*Yes*	*Yes*	*Yes*	*Yes*
Observations	3472	3472	3462	3437
Adjusted R^2	0.035	0.010	0.004	0.002

Notes: Coefficients and robust standard errors are reported for OLS regressions of daily ticket orders on purchase-date sunshine duration (in percent of time), purchase-date rainfall (in 1/100 mm), and horizon indicators (dummy variables for the number of days between purchase-date and movie-date). Fixed effects for the show are included. One observation is the number of sales for a particular show per day. The column headings indicate how many days in advance tickets are purchased in the sample used.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

can be explained by projection bias (or a reminder-effect of good weather), we expect to find an effect on early orders as well.

To analyse this prediction, we estimate (a variant of) the fixed effects model (1.4) for separate sets of advance sales which are defined by how many days in advance tickets were sold. More precisely, we estimate the effect of weather on ticket orders between 5 and 11 days in advance. Since weather forecasts for this time horizon are lacking, we cannot include the term $\mathbf{F}_{\tau\mathbf{t}}\beta_{\mathbf{F}}$ in these regressions. We repeat this exercise for time spans between 9 and 15, 13 and 19, and 17 and 23 days in advance.[26]

Table 1.5 reports the results. It becomes apparent that the effect of average sunshine duration on sales is significantly greater than zero at least at the five percent level for most estimated models. The notable exception are the results with sales between 13 and 17 days in advance; here, the coefficient of sunshine duration is rather small.[27] Regarding the economic interpretation of the estimates, one standard deviation change in weather explains variations in sales between 10 and 17 percent of (quite low) mean sales for the respective periods and is therefore in a similar range as for all our estimates in Section 1.4.1.

[26] Obviously, the choice of beginning and end days of these time-spans is arbitrary. However, our qualitative results do not depend on the exact location of the time spans as long as they are sufficiently long (greater than four days) to allow for enough within variation for early sales.

[27] We observe fairly small coefficients for all intervals which include the time horizon of exactly 16 days. Excluding observations for this time horizon leads to significant coefficients throughout, which suggests that this time horizon is an outlier for which we have no plausible explanation.

Table 1.6: Effect of Hourly Changes in Weather on Changes in Ticket Orders

	Hourly Ticket Orders		
	Morning & Afternoon	Morning	Afternoon
Diff. Sun per h.	0.00038**	0.00043	0.00043*
	(0.00018)	(0.00029)	(0.00024)
Diff. Rain per h.	0.000019	−0.000028	0.000027
	(0.000038)	(0.000042)	(0.000048)
Hour Indicators	*Yes*	*Yes*	*Yes*
Observations	22454	8984	11225
Adjusted R^2	0.001	0.001	0.000

Notes: We report the coefficients and robust standard errors of OLS regressions of the first difference of hourly ticket orders on the first difference of hourly purchase-date sunshine duration (in percent of time), the difference in purchase-date rainfall (in 1/100 mm), and hour indicator variables. Column 1 reports coefficients for all orders between 8 am and 8 pm. The two remaining columns split the dataset into orders before and after 2 pm.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Note that these results can also be seen as a robustness check for the concern that information content of current weather drives the results. As early as three weeks before the movie-date, (perceived) information content of purchase-date weather for movie-date weather should be nil.

Another instance in which changes in weather can naturally be assumed to have little impact on the probability of ticket availability are variations in weather from one hour to the next. Again, we only expect to find an effect of hourly changes in weather on changes in ticket orders if the state of the world by itself – and not its effect on results of market interactions – affects choice behaviour.

To test this prediction, we regress, for a given movie-date, the first difference of ticket orders per hour on the first difference of sunshine duration and precipitation. In addition, hour dummies are included as independent variables to control for different sales volumes over the course of the day. We restrict the sample to hours with potentially positive sunshine duration (8 am to 8 pm) as well as to hours in the morning (8 am to 2 pm) and afternoon (2 pm to 8pm) between one and four days ahead of the movie-date.

In this analysis, the effect of weather on ticket orders is identified through variations in weather within a given day. Given the low within-day variation in sales and thus low mean differences as dependent variable, the estimated coefficients are rather small (see Table 1.6). Still, hourly changes in weather have a statistically significant effect on changes in ticket orders. The estimates seem to be mainly driven by sales in the afternoon when most tickets are ordered and therefore hourly variation in sales is highest.

In light of the evidence that both, very early sales as well as hourly changes in sales are affected by current weather, we conclude that current weather explains ticket order at which the probability of ticket availability is independent of purchase-date weather. Summarizing our above arguments, projection bias is the only explanation, which can simultaneously account for all our empirical findings.

1.5.3 Robustness

We examine the robustness of our empirical results along various dimensions.[28] First, we check whether the results depend on estimating linear models in the form of (1.4) and (1.5). We therefore repeat the entire analysis using count data models (poisson and negative binomial regressions). None of the results are altered by estimating either of these models.

Second, we evaluate whether the main results in Table 1.3, for which we have most statistical power, are driven by specific subgroups of customers, movie genres, or the weekday of the show. To this end, we repeat the analysis for the subgroup of sales conditioning on the timing of the order (morning, afternoon, and evening), the number of tickets ordered (one, two, three, and more than three), the age of the buyer (below and above 30 years), movie genre (drama, comedy, and action/adventure), and the weekday of sale and movie respectively (weekend, weekday). The number of total orders across these subgroups varies considerably – for example, more than 50 percent of ticket orders comprise of two tickets, while less than 10 percent of customers order three tickets – which is accompanied by different degrees of statistical power for the analysis.

Nevertheless, the estimated effect of sunshine duration and (to a lesser extent) rainfall on total orders continues to be statistically and economically significant at least for the fixed effect model (1.4). The only exception from this rule are ticket orders on weekends (Friday to Sunday), for which the number of observations is lowest (N = 291). Additionally, the cross-section models (1.5) yield meaningful effects for current weather on sales for the majority of specifications.

Third, we check whether our results are driven by the selection of average sunshine duration as the relevant weather variable. To do so, we substitute average sunshine duration by average temperature in all regressions with ticket orders as the dependent variable. By and large, this substitution leaves the results unchanged. Similarly, we test whether our conclusions are sensitive to the choice of the independent variable. Instead of using

[28]Due to the large number of robustness checks, we do not include the detailed results in this paper. The estimation results for all analyses mentioned in this subsection are available from the authors on request.

aggregated ticket orders as the quantity to be explained, we could have also used the overall number of tickets sold as independent variable. These two measures are highly correlated, such that it is not surprising that this modification does not lead to different conclusions.

A fourth possible concern could be that ticket sales depend on recent rather than current weather. For example, customers may be more inclined to buy tickets if weather was good for a couple of days, possibly indicating a stable high pressure weather system. We account for this by including one period lagged weather indicators in our analysis of models (1.4) and (1.5). In these models, current weather persists to have explanatory power, contrary to weather one day earlier.

Finally, we examine the robustness of all empirical results to the inclusion and exclusion of various control variables. First, we attempt to proxy for the probability of ticket availability directly by (a) including a dummy indicating whether the theatre turned out to sell out for the particular show of interest and (b) including the number of ticket orders until the purchase-date in all variants of cross-section models (1.5). Second, we control for the popularity of the movie by including either the number of theatres in which the movie was shown on the opening weekend in Germany, or movie gross in Germany on the opening weekend (or both) as independent variables in all variants of model (1.5).[29] Neither of these modifications changes the results in any meaningful way. The same holds true for excluding controls $\mathbf{D}_{\tau\mathbf{t}}$ and $\mathbf{X}_{\tau\mathbf{t}}$, respectively, in all the models we estimate.

1.6 Conclusion

There is a growing literature which shows that projection bias impedes the ability of individuals to consistently predict future utility. Predicted utility at unknown future states of the world tends to be biased towards utility at today's state of the world.

This paper presents evidence for projection bias in a simple decision problem (purchasing advance tickets for an outdoor movie theatre), in which the transient nature of today's state is obvious due to risk and explicitly pointed out to decision makers. The availability of unbiased and precise forecasts regarding future states, on which individuals may condition their decisions, should further reduce the extent to which projection bias affects

[29] The choice of the popularity indicator is somewhat arbitrary; for example, we could have chosen the total gross of the movie shown as well. We opted for opening weekend measures to avoid measurement error due to the total time the particular movie has been screened in theatres. The concern for measurement error arises, because the outdoor movie theatre in question shows recent films as well as classics. All data for this analysis have been retrieved from the database http://www.boxofficemojo.com [October 2011].

choices. We find that the current state – current weather – influences choices to a large extent, which suggests that de-biasing decision makers may turn out to be challenging.

Put in a broader context, our result that projection is present even in very simple decision problems points towards the possibility that it may have important aggregate implications, specifically when decisions of individuals observing the same state of the world are biased in the same direction. It has been shown recently that such correlated errors can be amplified through feedback effects in markets, leading to potentially large fluctuations (Hassan and Mertens, 2011). For this to be the case, further research is needed to answer the question whether projection bias affects choices over alternatives, whose utility depends on an endogenous state of the world (like consumption and savings decisions depending on the state of the economy).

Finally, studying projection bias under risk highlights the need to understand how exactly individuals mispredict future utility. Do they indeed undervalue the extent to which utility varies with the state (and hold correct beliefs regarding future outcomes), the standard interpretation of projection bias? Or are their beliefs regarding the likelihood of future states biased towards the current state (and the predictions of state-dependent utilities correct)? Answering these question is certainly important for finding ways to help individuals to predict future utility accurately and to make good decisions.

Chapter 2

Reciprocity in Organisations – Evidence from the WERS[0]

2.1 Introduction

Understanding the behaviour of employees in labour relations is crucial for managers and firm owner who aim to align potentially diverging interests of management and workforce. In the last decades contract theorists developed a consistent framework in which monetary incentives induce agents to exert effort, serving as a guideline for real-world firms.[2] While modern human resource departments (to some extent) rely on theoretical considerations, the majority of real-world labour contracts are characterised by fixed payments and – if at all – only a minor part of employees' income is attributed to incentive pay.[3] The classic static moral hazard theory would predict lower levels of effort exertion than the real-world examples show.

Incorporating concepts from behavioural economics may provide additional explanations for real-world observations. In an early contribution, Akerlof (1982) demonstrates that wages may exceed the market-clearing wage when employers attempt to influence working norms via gift-exchange.[4] More recently, Englmaier and Leider (2012) introduce the concept of reciprocity – i.e. gift-exchange motivation – into the classical principal-agent framework concluding that firms with reciprocal employees have more leeway to cost-efficiently induce effort: shifting away from direct monetary incentives and induc-

[0]This chapter is based on joint work with Florian Englmaier and Stephen Leider.

[2]See Prendergast (1999) for a survey.

[3]Lemieux et al. (2009) estimate that approximately 37% of male labour market participants in the US (using the PSID, Panel Study of Income Dynamics 1976 - 1998) receive variable payments with a median magnitude of 3.5%. Englmaier and Leider (2012) discuss further studies corroborating this argument.

[4]Also in a labour market context, Becker et al. (2011) provide field evidence for heterogeneous long-term responses to gift-exchange motivation.

ing employees to behave reciprocally towards their employers allows firms to save high costs from risk premia they would have to pay when using strong incentives. Relying on reciprocity, however, requires firms to screen for employees with reciprocal traits.

In this paper we use Englmaier and Leider (2012) as a theoretical guideline and search for evidence for the use of reciprocity based motivation in organisations. Using the 5th wave of the "Workplace Employment Relations Survey" (WERS 2004) a large scale survey of Britain-based firms, we find evidence for firm behaviour consistent with gift-exchange motivations. We use compulsory personality tests for job candidates as an indicator whether firms explicitly screen applicants for personality traits that may be correlated with job candidates' inclination towards reciprocity. In line with gift-exchange motives, these firms are more likely to provide their employees high wages and other non-pecuniary benefits like employer pension schemes and extended paid annual leave. Furthermore, employees in these establishments enjoy more on-the-job training (c.f. Leuven et al. (2005)) and have a higher chance that their employer provides guaranteed job security.

Screening applicants' personality and providing benefits for those who get hired may pay for the firm if employees reciprocate with higher effort. Even though we are not able to measure effort directly, we find that employers using personality tests report higher levels of firm performance and are more likely to organise work in teams. The latter is particularly interesting as reaping synergy effects when working in teams requires team members to subordinate their own desires to the common good.

In contrast, two additional measures of modern human resource practises – competency tests for job candidates and variable payments for employees – fare in general much worse in predicting benefits for employer or employees (if they do at all). This implies that only screening for personality as opposed to the use of competency tests or other human resource practises explains patterns consistent with gift-exchange motives. The lack of an association between personality tests and dismissals within an establishment furthermore provides suggestive evidence that personality tests do not merely increase the "fit" between employer and employee, which otherwise might have caused similar relationships between screening and benefits.

Closest to this work is Huang and Cappelli (2010). Based on a national survey of US employers they argue that employers who state that they particularly value applicants with high "work ethic" are less prone to monitor their employees, organise more work in teams, and have lower turnover rates. Furthermore employees receive higher wages and firms are more productive. Comparing their results with ours we can, by and large, confirm their findings, with the exception that we do not find any relationship between personality tests and turnover and monitoring respectively.

Despite some similarities with Huang and Cappelli (2010), we are distinct in at least two main dimensions: First, we are able to go beyond the analysis in Huang and Cappelli (2010) as the richness of our data allows us to include the entire range of occupational groups within an establishment, from managers to unskilled labour, into our analysis. Huang and Cappelli (2010) are restricted to data on frontline worker only. Both studies, however, have in common that the cross-sectional structure of the data does not allow to pin down a unique explanation of the observed pattern. Despite evidence in favour of reciprocity as the underlying principle we cannot establish causality of the effects.

Second, Huang and Cappelli (2010) use a survey question in which managers have to rate how important candidates' "work ethic" is for them when assessing applicants. In contrast, we use hard information on whether written personality tests are used in the hiring process. These tests are based on observable practise, implying that other datasets may contain this measure as well which ensures that the analysis is transferable to other data sources containing information on test use.

Personality tests are only one potential dimension of how firms screen job candidates. Among other popular methods are interviews, reference letters and – widely used – competency tests.[5] Whereas the latter aims to uncover cognitive ability, personality tests – the "Big Five" framework is a prominent example – measure a whole range of characteristics of a potential employee. In particular we interpret the use of these personality tests as a proxy for firms that are more likely to have (highly) reciprocal workers – in most cases due to screening also for other desirable traits that are correlated with reciprocity, though in some cases firms may be directly screening for reciprocity.

Our empirical approach is consistent with findings that document that personality traits usually identified with personality tests within the "Big Five" framework are (closely) correlated with measures of *reciprocity* as commonly defined in laboratory experiments.[6] Ben-Ner et al. (2004a) link behaviour in a dictator game with switching roles to previously elicited personality traits and find that "Big Five" indicators "agreeableness" and "openness" are associated with higher amounts a dictator sends in response to the amount she previously received. Opposed to that, cognitive ability seems not to influence the propensity to reciprocate. In an earlier contribution, Ashton et al. (1998) concludes on basis of hypothetical questions that high "agreeableness" and high "emotional stability" are associated with high reciprocal altruism. Consistent with the patterns in our data,

[5]See Rynes and Cable (2003) for an extensive review on the various methods employed in modern hiring procedures.

[6]Autor and Scarborough (2008) document the hiring procedures of a large retail firm, which (according to the authors) is representative for the industry, that uses personality tests to screen workers upon hiring. The firm gave hiring preference to applicants with positive z-scores for "agreeableness", "conscientiousness", and "extroversion", "Big Five" traits that are predictive for the presence of reciprocity.

Autor and Scarborough (2008) provide evidence that firms widely make use of screening for personality and Wilk and Cappelli (2003) show that employers differ substantially in the extent to which they make use of applicant screening.

The importance of social preferences for individual decisions has been documented in various studies, for extensive surveys see Fehr and Schmidt (2003) and for field evidence DellaVigna (2009). Fehr and Gächter (2000) in their survey explicitly concentrate on the prevalence of reciprocity. In several theoretical contributions, social preferences have been associated to optimal contract designs, suggesting that not only productivity and ability but also social traits can influence the generosity of contract offers.[7] In an empirical study using survey data, Dohmen et al. (2009) provide evidence from real-world labour markets for the importance of reciprocity on wages and effort provision. Englmaier et al. (2011) in a real-effort laboratory experiment elicit both productivity and social preferences from agents and find that principals increase wages for both traits by adapting contract offers accordingly. In an earlier contribution, Cabrales et al. (2010) predict outcomes in a gift-exchange experiment on basis of elicited behavioural preferences.

Another strand of the personnel literature explores synergies between different human resource practises. Using firm data from steel finishing lines, Ichniowski and Shaw (1997) find that modern human resource practises are associated with productivity of these firms. More detailed, practises like incentive payments, work organised in teams, flexible job assignment, job security, and training for the employees has positive effects on productivity. In a recent experimental study Bartling et al. (2012a) find complementarities between high discretion, high wages and rent sharing, job characteristics which are commonly associated with "good jobs" (see Part III of this dissertation). The authors demonstrate that these jobs emerge endogenously (as they are profitable) if employers have the opportunity to screen job candidates. Importantly, they show that it is screening for social preferences and not for competency which is necessary for "good jobs" to emerge.

We contribute to the literature by combining evidence from both strands of the literature. Using field data, we show that screening for personality when employers hire new employees is associated with a bundle of benefits for employees and employers like on-the-job training, workflows organised in teams, provision of additional (possibly non-pecuniary) benefits, or better performance of the firm per se. In contrast to these findings, competency tests predict outcomes by far worse. Moreover, personality tests are unrelated to dismissals within firms. Hence, explanations solely targeting on correlations between successful firms and application of modern human resource practises are too narrow and we feel confident that our results point at a more nuanced, behavioural, explanation of

[7] See for references Itoh (2004), Dur and Glazer (2007) or Englmaier and Wambach (2010)

the observed patterns: The systematic use of reciprocity based motivation by firms.

The remainder of the paper is structured as follows: In Section 2, we provide an extensive description of the WERS 2004 with a special focus on personality tests. Section 3 contains details on the Hypotheses, the estimation strategy, and results. Furthermore a substantial part of the section is dedicated to robustness checks. Section 4 concludes with a discussion.

2.2 Data

2.2.1 The WERS 2004 Dataset

The empirical analysis relies on the 2004 "Workplace Employee Relations Survey" (WERS 2004)[8], the fifth in a government-funded series of surveys carried out at British workplaces. The WERS 2004 covers information on employment relations of British workplaces and is provided by employees and employers. The following analysis entirely relies on the information about establishments provided by employers.[9]

The WERS 2004, consisting of 2,295 establishments surveyed, is a representative sample of the British economy.[10] The number of employees per establishment varies widely between a minimum of 5 jobs per workplace up to 10,006 with an average of 414 jobs per workplace. Note, however, that the mean is inflated by few extremely large companies – the median firm size is 69 jobs and even the 99th percentile only contains a maximum of 4,936 jobs per workplace. The firms cover almost all branches of the economy with a slight concentration on health, whole trade and manufacturing.[11] About one fourth of the establishments are attributed to the public sector. More than half of the establishments are unionised (58 percent).

21 percent of the establishments are part of the productive sector, three quarters are one of a number of different workplaces in Great Britain belonging to the same organisation, 23 percent are a single independent establishment and 2 percent are the sole UK establishment of a foreign company. Overall 78 percent of the firms are either en-

[8]For further information on the WERS see: http://www.wers2004.info.

[9]The WERS consists of different datasets with varying respondents. Besides the survey answered by employers, the WERS also comprises datasets on employees and employee representatives with questions targeted to figure out their individual view on the establishment and their working conditions. The latter two datasets are not employed for our analysis.

[10]"WERS 2004 (...) provide(s) a nationally representative account of the state of employment relations and working life inside British workplaces." Source: http://www.wers2004.info/wers2004/wers2004.php, October 23rd 2012.

[11]Workplaces are classified according to the SIC 2003 (Standard Industrial Classification) by the UK National Statistics. Sectors not covered by the WERS 2004 include: Agriculture, hunting and forestry, fishing, mining and quarrying, private households with employed persons, and extra-territorial bodies.

tirely or predominantly UK-owned, whereas the controlling head office of the company is foreign-based in only 12 percent of the cases. Market shares are widely dispersed with approximately 39 percent (15 percent) of the firms indicating a market share of less than five percent (more than 50 percent). Roughly in line with this, about 75 percent of the firms report that the perceived degree of competition in their market is either high or very high whereas 11 percent state it to be low or very low.[12]

Within each firm the WERS 2004 distinguishes between 9 different occupational groups.[13] Panel (d) of Figure 2.1 provides absolute frequencies for all nine occupational groups pooling all 2,295 establishments. Not surprisingly, almost all firms state to have a management department and about 80 percent of the surveyed firms have employees in secretarial or administrative positions. As several variables of interest, including modern human resource practises, are provided on occupational group level, our subsequent analysis relies on both, firm level and occupational group level.

Table 2.1 provides summary statistics for variables of interest, including the following statistics: the number of observations, averages and standard deviations, the 25th, 50th and 75th percentile as well as minimum and maximum values. The first set of variables is reported on firm level.

"Monitoring" is an ordinal variable asking for the proportion of non-managerial employees who have job duties which involve supervising other employees. Value one indicates that no employee has monitoring tasks – on average firms indicate that between one and 19 percent of the workers have monitoring tasks. The continuous variable (relative) "Dismissal" measures the percentage of the workforce which has been dismissed within the previous year. The data suggest that dismissals occur very rarely.

"Firm Performance" is an indicator which combines the following self-reported performance measures: "Financial Performance", "Labour Productivity" and "Product Quality". The indicator is one, if firms in at least one of the three dimension report to have better performance than the median answer of all firms for each dimension.[14] This clas-

[12]The fractions of the legal state, market share and the degree of competition are calculated dropping any unclear answers.

[13]These occupational groups are: (1) Managers and senior officials, (2) professional occupations, (3) associate professionals and technical occupations, (4) administrative and secretarial occupations, (5) skilled trade occupations, (6) caring, leisure and other personal services, (7) sales and customer service occupations, (8) process, plant, and machines operatives, and drivers, and (9) routine and unskilled occupations.

[14]For these three measures employers were asked to rate the performance of their firm compared to the relevant industry, resulting in heavily over-rated own performance: for the example of "Labour Productivity", 49 percent of employers state to be better or a lot better than the average and 94 percent state to be at least about average for the industry. Overrating is similarly severe for variables "Financial Performance" and "Product Quality". To account for this overrating we classify establishments as successful if their own rating is better than the median rating of all firms.

Table 2.1: Summary Statistics

	Obs.	Avg.	SD	Pctl. 25	50	75	Min.	Max.
	Firm Level							
Monitoring	2278	2.52	1.08	2	2	3	1	7
Dismissal	2159	0.01	0.04	0	0	0.01	0	0.88
Firm Performance	2160	0.54	0.5	0	1	1	0	1
Firm Benefit	2295	0.8	0.4	1	1	1	0	1
Firm Benefit 2	2295	0.89	0.31	1	1	1	0	1
Job Security	2295	0.17	0.37	0	0	0	0	1
Top Wage	2135	0.18	0.24	0	0.08	0.27	0	1
Low Wage	2135	0.03	0.13	0	0	0	0	1
Personality Tests	2292	0.34	0.47	0	0	1	0	1
Competency Tests	2291	0.61	0.49	0	1	1	0	1
Incentive Pay	2295	0.57	0.49	0	1	1	0	1
	Largest Occupational Group							
Possibly Non-Pecuniary Benefits								
Any Benefit	2286	0.89	0.31	1	1	1	0	1
No. Benefis	2286	2.6	1.36	2	3	3	0	5
Pension Scheme	2286	0.77	0.42	1	1	1	0	1
Company Car	2286	0.18	0.38	0	0	0	0	1
Private Health	2286	0.20	0.40	0	0	0	0	1
Extended Paid Leave	2286	0.75	0.43	1	1	1	0	1
Sick Pay	2286	0.70	0.46	0	1	1	0	1
On-the-Job Training	1950	4.05	1.09	3	4	5	1	6
General Training	2288	0.58	0.49	0	1	1	0	1
Team-working	2279	5.08	2.25	3	6	7	1	7

Notes: Statistics for each variable are calculated omitting "refusal", "don't know" and "not applicable", indicating unclear answers. "Job Security", "Personality Tests", "Competency Tests" and "Incentive Payments" are collapsed on firm level to guarantee comparability. The lower panel refers to information on the largest occupational group only. "Monitoring", "On-the-Job training", and "Team-Working" are ordinal variables with lower values corresponding to lower levels of monitoring, training, and team-working respectively. "Dismissal", "Top Wage", and "Bottom Wage" are continuous fraction of dismissed employees, and employees with high and low earnings. "No. Benefits" counts the number of granted benefits, the remaining variables are binary.

sification based on self-reporting, splits the data into two almost equal parts of rather successful and unsuccessful establishments.

We construct the variable "Firm Benefit" as a comprehensive measure of successfulness of establishments. It either relates to self-reported outcomes or to the ability of the firm to employ highly desirable work practices: The measure takes the value one if the respective firm either reports higher than median firm performance, uses more team-working than the median firm, or relies less on monitoring (compared to the median).

We also use an alternative indicator for overall firm benefits, “Firm Benefit 2” which includes dismissals and reports high benefits if additionally the firm has dismissals lower or equal to median dismissals. The purpose of this procedure is to fully address all firm benefits – monitoring, team-working, dismissals, and productivity – which were suggested by Huang and Cappelli (2010) in one compound measure. However this procedure comes at a price: By doing so, we lose much of the variation as most firms do not have any dismissals within the previous year, c.f. Table (1), resulting in almost 90 percent of firms being classified as firms which reap some suggested benefits.

The data contain rich information on various aspects of the workers compensation and benefit package. “Job Security” is reported for each occupational group. In this table, however, we collapse the measure on firm level. Hence the dummy variable on “Job Security” is one if at least in one occupational group within an establishment employees enjoy job security or non-compulsory redundancies.[15] Finally, “Top Wage” (“Low Wage”) is an indicator variable which provides information about the relative size of top-wage (low-wage) earners compared to all employees within a firm. The WERS defines the highest wage category (we label this category “Top Wage”) as wages equal or more than 15 pounds per hour. The dataset provides three more wage categories: 4.5 or below (“Low Wage” category), 4.51 – 5, 5.01 – 15 or 15 and above pounds per hour.

For the second set of variables the dataset provides measures for the largest occupational group in terms of employees. Possibly non-pecuniary benefits for the employee comprise different measures of benefits for a worker. “Any Benefit” is a binary variable indicating whether employees of the largest occupational group receive any of five benefits suggested in WERS 2004, with almost 90 percent of firms providing at least one benefit.[16] “Number of Benefits” is an ordinal measure how many (between zero and five) different benefits of the suggested five benefits employees receive. We furthermore provide summary statistics for all in the survey suggested benefits, namely “Pension Scheme”, “Company Car”, “Private Health Insurance”, “Extended Paid Leave” and “Sick Pay”.

The variable “On-the-Job Training” is measured ordinally with value one indicating that employees of the largest occupational group did not experience any training within the previous year and six implying ten days and more. The WERS 2004 furthermore distinguishes between providing training on computing skills, team-working, communicational

[15]Non-compulsory redundancies cover voluntary redundancies and early retirement, see https://www.gov.uk/staff-redundant/noncompulsory-redundancy, November 20, 2012.

[16]The survey asks for the following non-pay terms and conditions: (1) Employer pension scheme, (2) company car or car allowance, (3) private health insurance, (4) more than four weeks paid annual leave, (5) sick pay in excess of statutory requirements, and (6) none of these.

skills, leadership skills, operation of new equipment, customer service, health and safety, problem-solving methods, equal opportunities, reliability and working to deadlines and quality control procedures. We classify team-working skills, communicational skills, and leadership skills under the label “General Training” as these cover matters which are not narrowly job-specific but may be considered as more general and hence can be beneficial for an employee’s entire working life across different employers.

Finally “Team-working” is an ordinal variable asking for the proportion of employees in the largest occupational group being designated to teams. No team-working at all (value one) is rather rare, and an average of almost three indicates that 60 – 80 percent of largest occupational group employees work in teams.

2.2.2 Modern Human Resource Practises

WERS 2004 provides detailed information about human resource practises within establishments including the prevalence of personality tests, competency tests and variable payments by occupational group. In order to be able to control for the fact that some workplaces might have an in general more sophisticated HR department we classify establishments that use these three practices as employing modern human resource practises.

More than one third of all establishments use personality tests when screening job candidates whereas more than 60 percent of firms make use of competency tests in at least one occupational group (see Table 2.1). Both personality tests and competency tests are less prevalent in sectors with lower skill intensive tasks (i.e. construction, wholesale and retail, and hotels and restaurants) while we find high rates of competency tests in financial services, public administration and education.[17] Similarly, personality tests are prominent in financial services, public administration and manufacturing.

Analysing the prevalence of both screening tests within the firm it is no surprise that screening devices are most common for hiring managers. Excluding managers, in about 24 percent (56 percent) of establishments personality (competency) tests are required at least in one occupational group when recruiting new employees.

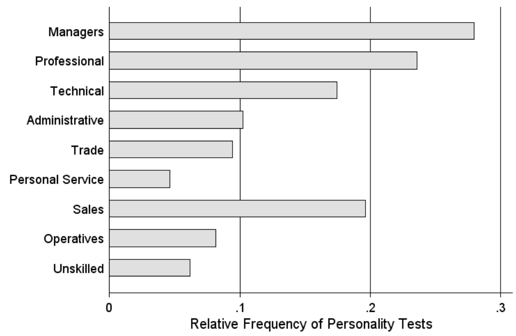
The prevalence of personality and competency tests by occupational group is summarised in detail in panel (a) and (b) in Figure 2.1. Comparing both panels it again becomes clear that employers use competency tests more often when hiring applicants: for each occupational group the relative frequency of competency tests exceeds that one of personality tests. More interestingly, the distributions of both tests differ to a large extent.

[17] One notable exception of lower skill intensive tasks and very high rates of both tests is the sector “electricity, gas and water” as classified by the UK National Statistics. However this may not be representative due to small sample size of only 45 observations for this sector.

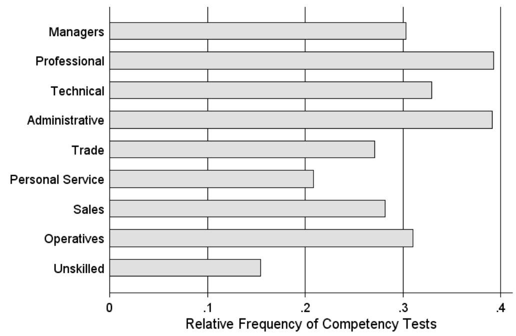
Figure 2.1: Relative Frequency of Human Resource Practises by Occupational Group

This figure provides an overview over the prevalence of personality tests, competency tests (upper row) and incentive pay for nine different occupational groups (bottom left). The picture on the bottom right depicts the prevalence of each of the nine occupational groups in absolute terms.

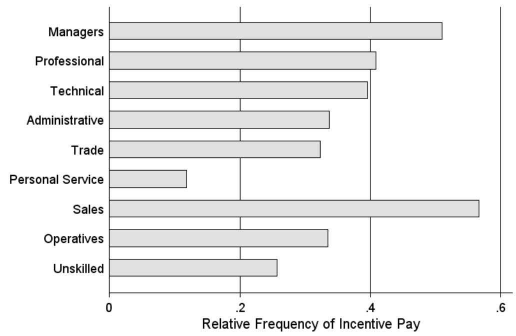
(a) Personality Tests



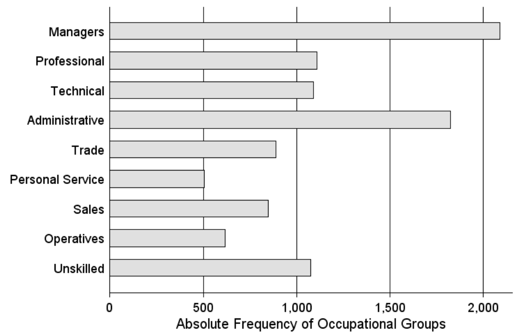
(b) Competency Tests



(c) Incentive Pay



(d) Occupational Group



Whereas firms make use of competency tests to rather similar extent across occupational groups (with exceptions of personal services and unskilled labour with clearly lower rates) the prevalence of personality tests starkly declines with decreasing skill intensity. The exception is the group of sales employees who are very likely to be screened for personality upon hiring.

A comparison of both distributions provides some tentative evidence that personality tests and competency tests are measuring different characteristics of the job candidate and are applied to different job requirements. This assessment is further supported by a correlation coefficient of only $\rho = 0.17$ between personality tests and competency tests for non-managers implying no strict path dependency in firms' choice of which screening devices to apply.[18] Of all firms, 38 percent only screen for competency and 5.5 percent exclusively screen for personality upon hiring, whereas 39 percent apply both devices.[19]

[18]This measure correlates personality tests and competency tests both for all occupational groups excluding managers.

[19]Including managers shifts these fractions a bit: 38 percent of firms use exclusively competency tests,

The third measure for advanced human resource practises are incentive components in employees' compensation schemes. Paying some sort of variable payment – either performance pay or profit pay – is common in 57 percent of establishments (and in half of the firms in at least one occupational group if abstracting from the group of managers).[20] As can be seen from panel (c) in Figure 2.1 the distribution of incentive pay across occupational groups declines for less skill intensive tasks with the exception of sales, where incentive pay is common.

2.3 Reciprocity in Organisations

In this Section we use the presence of personality tests in a firms' hiring procedure as a proxy for this firm having a weakly more reciprocally inclined workforce. Even if a firm does not use personality tests to directly screen for reciprocal workers, it might end up with a more reciprocal workforce as a by-product (see the discussion in the Introduction). This allows us to test various hypotheses regarding reciprocity in organisations within one data set. Some of these hypotheses have independently been advanced in Leuven et al. (2005), Huang and Cappelli (2010), and Englmaier and Leider (2012).

2.3.1 Hypotheses

The model in Englmaier and Leider (2012) serves as a loose theoretical background for developing the following hypotheses. In this model employers can employ incentives based on gift-exchange if two conditions are fulfilled: First, in a labour market with heterogeneous agents, the employer has to screen for reciprocal job candidates, willing to repay a generous contract offer with increased effort. Second, the willingness of these reciprocally inclined employees to reciprocate needs to be "activated" by the employer via initial "kind behaviour". More technically, the employer has to offer a contract that exceeds the agent's outside option. This can be achieved by offering a higher than market wage[21] or, as the model is based on utility arguments, by providing other, possibly non-pecuniary, benefits, like an employer pension scheme or paid annual leave.

According to Leuven et al. (2005), firms with a more reciprocal workforce are more likely to provide training to their employees. Besides regarding training as additional benefit for

7 percent personality tests and 39 percent both tests.

[20]The dataset only indicates whether a firm provides variable pay for a certain occupational group but does not give estimates of its magnitude compared to the fixed wage.

[21]A different explanation of high wages is provided by Huang and Cappelli (2010): As workers with high work ethic help the firm to save costs, employers attempt to hire as many of these types as possible, which drives up their wages (rent sharing).

workers, reciprocal behaviour of the agent might be a necessary condition for the provision of (excess) on-the-job training. As benefits from training are inherently sequential, the employer has to trust her employee that the employee does not enjoy the training and then leaves for a better offer. Put differently, training could be regarded as an increase in the worker's outside option. Furthermore if the employer is convinced of the worker's reciprocal behaviour, she may be willing to provide relatively more general training, which is advantageous not only for a specific job but for the worker's entire employment biography.

In a similar vein, employers may provide job security to their labour force, signalling confidence in workers' loyalty towards the firm. If agents, however, lack reciprocal attachment to the establishment, job security schemes enable employees to exploit this device via shirking while being protected against immediate consequences.

Hypothesis 2.1 (Generosity to Workers)**.** *Firms which screen for reciprocity pay higher wages, are more likely to provide their workforce additional (potentially non-pecuniary) benefits, should have a higher likelihood to provide their workers higher amounts of on-the-job training, in particular more general training, and should be more inclined to provide job security to their employees.*

On the other hand, making use of motivational devices which are based on reciprocal behaviour is costly for firms in the first place, as most gifts like higher wages or pension systems involve direct costs. Job security, for instance, inhibits employers to adjust the size of the labour force to fluctuations in demand in the short run. Hence, a rational employer using reciprocal motivation should expect to enjoy some benefits which at least offset these investments. Though the data does not allow us to pin down the cost-efficiency of a firm's behaviour, we proceed in our analysis by providing some insightful correlations.

Screening job candidates for their personality may be associated with employers' inclination to organise tasks in teams. If firms benefit from team-working under the condition of non-shirking and it is harder to measure effort of each team member compared to individual production (as the employer may only observe team output) then the implementation of team structures should be more likely in organisations with more reciprocal employees. This leads us to the hypothesis that organisations with compulsory personality tests and team-working of employees should be complements.

The strongest link between reciprocity and benefits for the firm are correlations of firm performance and screening job candidates for personality. Such relationships could imply that firms relying on reciprocity as a means of motivating workers on average are more successful in the market.

Huang and Cappelli (2010) document correlations between "work ethic" monitoring and turnover respectively. First, they argue that screening for "work ethic" and monitoring should be substitutes as employees with high "work ethic" exert effort voluntarily. Second, turnover decreases because the fit between job candidate and the firm should be better – a classical matching argument. For completeness, we likewise include these two results from Huang and Cappelli (2010) into the set of our hypotheses.

Hypothesis 2.2 (Value to Firms)**.** *Firms which screen for reciprocity should have more leeway to organise tasks in teams and should perform better in the market. According to Huang and Cappelli (2010) these firms could reduce monitoring and should have less turnover. In any case, firms should benefit in at least one of these dimensions when screening job candidates.*

2.3.2 Empirical Analysis

To study the effect of reciprocity on different outcome variables we use personality tests upon hiring as a measure for reciprocity. The general specification of our estimations is the following reduced form model:

$$y_{id} = P_{id}\beta_P + \mathbf{I'_{id}}\beta_\mathbf{I} + \mathbf{X'_i}\beta_\mathbf{X} + \epsilon_{id}$$

where y_{id} is the outcome of the dependent variable in occupational group d of firm i. The subscripts of P_{id}, an indicator for the use of personality tests, are defined accordingly. $\mathbf{I_{id}}$ are indicators, which are (for each establishment) available on occupational group level. $\mathbf{X_i}$ are (firm-wide) firm fixed controls and ϵ_{id} is an error term, which is assumed to be i.i.d. across firms but may be arbitrarily correlated within firms (between occupational groups). These potential within-firm correlations are accounted by clustering on firm-level.

Estimations differ in two main dimensions: First, we distinguish whether the dependent variable is reported at the firm-level or separately for each occupational group. The latter allows for matching between human resource practises and dependent variables on occupational group level. Second, different y_{id} are scaled differently, suggesting to adapt estimation strategies accordingly.

Job security – i.e. non-compulsory redundancies as defined in footnote 15 – is the only outcome variable which is provided for each occupational group; hence we estimate the effect of personality tests on job security pooling all available occupational groups.[22] This

[22] For this estimation we use all occupational groups per firm and create an indicator, whether the firm provides job security for employees in the respective occupational group. As differently sized firms may have more of less occupational groups (and hence giving firms with more groups a higher number of observations) and different sized firms may at the same time be differently likely to provide their

implies that we are able to match the provision of job security for each occupational group with the employed set of modern human resource practises.

The next set of dependent variables only contains information on the largest occupational group within an establishment. This set consists of all non-pecuniary benefits for the employee, including "Pension Scheme" and "Extended Leave", the "No. Benefits" as well as its prevalence ("Any Benefits"). Furthermore it comprises "On-the-Job Training", "General Training" and "Team-working". We adapt the model accordingly and replace the dependent variable with the outcome for the largest group $y_{i\widetilde{d_i}}$, $\widetilde{d_i} = max(\#\text{ of employees}(d_i))\ \forall i$ and $d \in \{$professionals, ..., unskilled occupations$\}$. We proceed analogously for personality tests $P_{i\widetilde{d_i}}$ and $\mathbf{I_{i\widetilde{d_i}}}$. Firm fixed controls which are summarised in $\mathbf{X_i}$ are unaffected.

For the remainder of the dependent variables, i.e. "Dismissal", "Monitoring", "Low Wage", "High Wage", "Firm Performance", and "Firm Benefit" the dataset only provides information at the firm level and lacks individualised occupational group specific data. Hence we construct aggregate measures from the occupational specific measure of personality tests, defining indicator P_i being one if in at least one occupational group (excluding managers and senior officials) job candidates are screened via personality tests.[23] Analogously to personality tests, we collapse $\mathbf{I_{id}}$ to the firm level and obtain $\mathbf{I_i}$.

Secondly, as outcomes are reported on different scales for different variables we adapt estimators accordingly. "Low Wage", "High Wage" and "Dismissal" are continuous variables, suggesting OLS estimation. Both employee benefits "Pension Scheme" and "Extended Leave" as well as the indicator whether the firm pays any benefits ("Any Benefit") are binary outcomes, implying probit regressions. The same applies to the variables "Job Security", "General Training", "Firm Performance", and "Firm Benefit". Finally "Team-working", "On-the-Job Training", "No. Benefits", and "Monitoring" are provided on an ordinal scale which leads us to use an ordered probit estimation approach.

The first set of controls, $\mathbf{I_{id}}$, comprises competency tests and a compound measure, whether employees (i.e. non-managers) either receive performance payments or profit payments. We define this measure as "Incentive Pay". These two variables can (along with personality tests) be regarded as indicators for modern human resource practises, which itself may be correlated with all outcome variables we observe. Controlling for them, we hope to reduce the problem of omitted variables.[24]

employees job security, we include the number of occupational groups per establishment into the set of controls. By this procedure we aim to reduce the likelihood that this effect may confound the results.

[23]The results are robust to the inclusion of managers. However, as the focus of this study is on reciprocal behaviour of employees, we exclude managers, who traditionally stand between the workforce and the owner of the company and hence may have different incentives.

[24]Note that we are fully aware of the difficulty to establish any causal effect of personality tests and

Table 2.2: Benefits for the Employee

	(1) OLS Bottom Wage	(2) OLS Top Wage	(3) O. Probit Training	(4) Probit Gen. Training	(5) Probit Job Security
Pers. Test	-0.051^{***}	-0.0066	0.26^{*}	0.34^{**}	0.22
	(0.014)	(0.015)	(0.14)	(0.17)	(0.14)
Comp. Test	-0.012	0.012	0.087	0.038	0.11
	(0.012)	(0.013)	(0.088)	(0.11)	(0.099)
Inc. Pay	-0.0059	0.049^{***}	0.14	0.13	0.042
	(0.014)	(0.012)	(0.097)	(0.12)	(0.12)
Foreign	-0.012	0.21^{***}	0.29	-0.29	-0.56
	(0.022)	(0.062)	(0.28)	(0.39)	(0.42)
Union	-0.034^{***}	-0.033^{**}	0.0061	-0.052	0.49^{***}
	(0.011)	(0.015)	(0.12)	(0.14)	(0.11)
PubSector	-0.043^{***}	0.00064	0.13	0.18	0.38^{**}
	(0.014)	(0.028)	(0.15)	(0.21)	(0.18)
Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2115	2189	1888	1964	7892
R^2	0.170	0.395			

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees within a firm, who earn less than 4.5 pounds per hour (“Bottom Wage”, reported in column (1)), the share of employees earning the “Top Wage” (more than 15 pounds per hour, reported in column(2)) as well as of probit regressions of provision of “Training” (column (3)), “General Training” (column (4)), and “Job Security” (column(5)), on dummy variables personality tests, competency tests, and on controls. Regressions in the first two columns provide results on firm level, and estimates for column (3) and (4) report estimates for the largest occupational group. Column (5) provides estimates for each occupational group and includes an additional dummy to control for the number of occupational groups per firm.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Firm fixed controls are summarised in $\mathbf{X_i}$, containing dummies for all nine possible occupational groups in a firm. We include these dummies whenever running regressions on occupational group level. In all regressions, we control for whether a firm belongs to a foreign organisation or is unionised. Furthermore we control for detailed recruiting practises and account for region, industry, size of the establishment, and use a dummy which indicates whether the establishment belongs to the public sector. As explained in footnote 22, estimating the effect of personality tests on “Job Security” includes the number of occupational groups into firm-level controls.

Table 2.2 and Table 2.3 summarise estimation results on Hypothesis 1, previously defined as necessary conditions to induce reciprocal behaviour of employees. We interpret personality tests as an indicator of whether employers search for potentially reciprocal workers.

we do not claim to be able to do so.

If they do so we should observe patterns associated with gift-exchange motivation.

Table 2.2 column (1) provides evidence that personality tests are significantly and negatively related to the relative size of employees receiving very low wages of 4.5 pounds per hour or less. Note also, that this is true for personality tests, but not for the other two proxies for modern human resource practises, competency tests and variable payments. In contrast, we do not find that personality tests can explain the relative share of employees who earn top wages, see column (2). These results provide some tentative evidence for the presence of a gift-exchange motive when employers screen on personality traits upon hiring.

Firms which screen their job candidates for personality offer significantly more days of on-the-job training per year (column (3)) and are more likely to train their employees with general skills which are beneficial for their future working life (column (4)). Similarly to low wages, neither competency tests nor incentive payments predict the amount of on-the-job training. The same applies to the matters that the training covers: the coefficient on competency test for instance is ten times smaller than the estimate of personality tests.

Finally, there seems to be a weak tendency for firms with obligatory personality tests to provide more job security, as shown in column (5). However, this relationship is not significantly different from zero. In Section 2.3.3 we repeat our analysis with different sets of human resource controls. There we find persistent and significant correlations of job security with personality tests.

Table 2.3 exploits information about non-pay terms and conditions in more detail. In column (1) we find personality tests being associated with the likelihood that at least one of five suggested benefits is provided by the employer.[25] The intuition for this result is simply that firms expect to receive some benefit from providing reciprocal incentives in terms of being able to use more progressive forms of production. However without data on the firm's costs for providing each single benefit we expect to see at least one of the possible benefits for firms using personality tests.

The next column provides evidence for a positive relation between personality tests and how many of five different benefits employees within a firm enjoy which confirms the results from the first column. Finally, we analyse two non-pay terms: the presence of an employer pension scheme and the provision of more than four weeks annual paid leave are strongly correlated with the use of personality tests.[26]

[25] The classification of these benefits are summarised in footnote 16.

[26] We report estimation results only for two out of five potential non-pay terms (c.f. footnote 16). These omitted conditions show systematically positive, though insignificant, correlations with personality tests. Estimation results are available from the authors upon request.

Table 2.3: Non-Pecuniary Benefits for the Employee

	Probit Benefits (1)	O. Probit No. Benefits (2)	Probit Pension (3)	Probit Extended Paid Leave (4)
Pers. Test	0.36*	0.29***	0.44***	0.36**
	(0.20)	(0.11)	(0.17)	(0.16)
Comp. Test	0.059	0.083	0.24**	0.13
	(0.13)	(0.078)	(0.11)	(0.11)
Inc. Pay	0.31**	0.36***	0.29***	0.34***
	(0.12)	(0.083)	(0.11)	(0.11)
Foreign	−0.35	0.0073	0.0029	−0.22
	(0.35)	(0.32)	(0.39)	(0.35)
Union	0.48**	0.39***	0.51***	0.52***
	(0.20)	(0.094)	(0.14)	(0.14)
PubSector	1.32***	0.43***	1.17***	0.83***
	(0.34)	(0.14)	(0.26)	(0.20)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275

Notes: We report the coefficients and robust standard errors of probit regressions of provision of "Benefits" (column (1)), provision of employer "Pension" scheme (column (3)), "Extended Paid Level" (column(4)) and ordered probit regressions of the "No. Benefits" (column (2)) on personality tests, competency tests, and on controls. All regressions provide estimates for the largest occupational group. Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

These benefits are not only strongly associated with personality tests but similarly closely related to variable payments. Note, however, that competency tests only poorly predict the provision of these benefits. Only for employer pension schemes (column (3)) we find significant correlations for competency tests, confirming our assumption that personality tests and competency tests are not substitutes.

Result 2.1 (Generosity to Workers). *Firms which screen for job candidates' personality are less likely to pay very low wages and provide more on-the-job training which then covers more general matters. Furthermore employees in firms with personality tests benefit from a higher likelihood to receive non-monetary benefits, especially employer pension schemes as well as extended paid leave and they receive a higher number of non-pay benefits overall.*

Rational employers provide gifts only if they expect to benefit from this strategy. Hence the second set of hypotheses is concerned with the employers' side.

Table 2.4 summarises potential benefits for the employer. Column (1) reports estimation results of personality tests on team-working. We find that the use of personality tests upon hiring is highly significant. Competency tests are associated to team-working as

Table 2.4: Benefits for the Employer

	(1) O. Probit Teamworking	(2) O. Probit Monitoring	(3) OLS Dismissal	(4) Probit Performance	(5) Probit Firm Benefit	(6) Probit Firm Benefit 2
Pers. Test	0.43***	0.081	0.0045	0.26*	0.51***	0.28*
	(0.15)	(0.10)	(0.0059)	(0.14)	(0.14)	(0.15)
Comp. Test	0.17*	−0.035	−0.00075	−0.089	0.076	0.0012
	(0.095)	(0.083)	(0.0041)	(0.099)	(0.11)	(0.14)
Inc. Pay	0.039	0.067	0.0076*	0.19*	0.083	−0.18
	(0.092)	(0.089)	(0.0043)	(0.11)	(0.12)	(0.15)
Foreign	−0.20	−0.39	−0.018***	0.12	0.33	0.18
	(0.25)	(0.29)	(0.0060)	(0.36)	(0.33)	(0.37)
Union	−0.032	−0.00094	−0.0087**	−0.15	−0.17	0.100
	(0.12)	(0.12)	(0.0036)	(0.12)	(0.13)	(0.13)
PubSector	0.23	0.048	−0.0065*	−0.058	0.22	−0.31
	(0.18)	(0.16)	(0.0038)	(0.19)	(0.24)	(0.23)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2268	2279	2149	2147	2279	2279
R^2			0.075			

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the degree of "Team-working" (column (1)), and "Monitoring" (column (2)), OLS regressions on the relative share of the variable "Dismissal" within one year in column (3) and probit regressions on firm "Performance" (column (4)) and in columns (5) and (6) two compound measures for overall "Firm Benefit" on dummy variables personality tests, competency tests, and on controls. Regression in the column (1) is based on the largest occupational group and column (2) – (6) provide results on firm level.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

well, but comparing the magnitudes of the coefficients, it becomes clear that the use of personality tests has more than twice the impact than the use of competency tests has. Furthermore it is important to notice that incentive pay has no influence on the likelihood of what fraction of employees are designated to teams.

The second two hypotheses are borrowed from Huang and Cappelli (2010). Contrary to their results, however, we do not find any relation between personality tests (or any other modern human resource practise) and monitoring. Similarly, there seems to be no (strong) relation between relative dismissals and any of the suggested human resource practises including personality tests.

At the same time the result on relative dismissals refutes the argument that firms do not use personality tests to primarily identify a specific trait, like social preferences and reciprocity, but are rather supposed to make sure an employee better "fits" to a company. If the latter was the case, we expect to see *fewer dismissals* in firms that use personality tests. Hence we conjecture that personality tests are not merely used to improve general fit of employees to firms but are devices to screen for social preferences.

Column (4) provides evidence that firms using personality tests seem to perform better, at least according to self-rated performance measures. Similarly to previous results competency tests have no explanatory power, whereas incentive pay predicts success comparably well.

Finally, columns (5) and (6) report results of the constructed measure on firm's benefits. "Firm Benefit" – i.e. whether work flows are organised in teams, the firms uses little monitoring, or reports high productivity – is highly related to personality tests and screening for personality is the only dimension of modern human resource practises which has predictive power. Furthermore, the alternative measure "Firm Benefits 2", which additionally to "Firm Benefits" assigns benefits to the firm if not a single employee was dismissed within the preceding year provides similar evidence.[27] For these two measures it is most striking that only personality tests can explain, whether firms are profiting in at least one of the suggested dimensions. However as personality tests and dismissals are unrelated, lower coefficients of "Firm Benefits 2" compared to "Firm Benefits" are not surprising.

Result 2.2 (Value to Firms)**.** *Firms which screen for job candidates' personalities designate more employees to work in teams and report to be more successful on the market. Pooling potential benefits, more team-working, less monitoring, better market performance (and less dismissals in a second specification) are highly related to the use of personality tests in hiring.*

2.3.3 Robustness

By providing robustness tests for the previous results, this section also offers an extensive discussion of our results so far. We are aware that drawing causal inferences is not valid as we cannot argue that personality tests were randomly assigned to firms. However, we go to great lengths to control for the general sophistication of a firm's human resource department. Modern human resource devices like personality tests or competency tests are likely to be correlated with (unobservable) other dimensions of quality of management practises which itself may be related to suggested benefits as well. Without being able to entirely exclude this mechanism, we aim to address that shortcoming by applying different sets of human resource practises as control variables.

We suggest five sets of human resource practises, only affecting the vector $\mathbf{I_{id}}$. Set 1 only includes whether the respective firm requires competency tests upon hiring and neglects incentive payments. Thus the indicator vector $\mathbf{I_i}$ only varies across firms, not within firms.

[27] See section 2.2 for a proper definition of "Firm Benefits 2".

The second set, Set 2, additionally includes whether the firm asks for personality tests *for managers*.[28] Set 3 additionally includes incentive payments. For the last two human resource sets we construct indicators reflecting potential complementarities between these measures: In Set 4 the indicator for modern human resource practises is equal to one if at least one of three, competency tests, personality tests for managers or incentive pay, is present at the respective firm. This measure has the least strict requirements for a firm to be classified as using modern human resource policies. In contrast to that, Set 5 requires firms to use all of the previously listed devices, implying it to be the strictest criterion for a classification in to the modern human resource category.

The appendix contains robustness tables on coefficients and standard errors for personality tests for each dependent variable and for each set of human resource controls. Note that even though not all coefficients of interest are significant at the highest level, the very systematic pattern of correlations emerging across a large set of specifications lends our core results substantial support. Table B.1 to Table B.5 refer to the table "Benefits for the Employee", Table 2.2. Both bottom and top wages are summarised in Table B.1 and Table B.2. "Bottom Wage" is related to personality tests for each set of controls, whereas "Top Wage" is not correlated to personality tests across any specification.

We observe similar behaviour of personality tests on employee benefits. Personality tests are significantly associated with "On-the-Job Training" in three of five control sets (Table B.3) and screening for personality of non-managers is significant in all regressions on "General Training" except when we explicitly include personality tests for managers (Table B.4).

In Table 2.2, "Job Security" is positively though insignificantly related to personality tests. Regarding Table B.5 we find significant association in four of five specifications. With weak evidence from our main regressions, we conclude to only provide some tentative evidence in favour of higher job security in establishments with personality tests.

The next set of tables, Table B.6 to Table B.9 relate to dependent variables in Table 2.3. Common to all four tables is that in specification 1, 4 and 5 coefficients change to only minor degrees and standard errors are comparable. Control sets 2 and 3 on the contrary depict smaller impacts of personality tests on dependent variables which in most cases – with the exception of employer pension scheme – lead to insignificant coefficients of personality tests. However, explicitly including personality test for managers (the decisive criterion of Set 2 and Set 3) into our analysis of whether personality tests influence suggested benefits changes the situation: Performing an adjusted Wald test for joint sig-

[28] Remember that managers are excluded in the entire analysis in order to avoid confounding results, as managers' job profiles involve both principal and agent duties.

nificance of personality tests for managers and personality tests for non-managers provides evidence for joint importance of personality tests. Personality tests for managers and non-managers are jointly significant at the ten percent level in Set 2 for “Any Benefit” (Table B.6). Both tests are jointly significant for both sets for “No. Benefits” on a one percent level, as can be seen in Table B.7. Finally, both tests are jointly significant for “Pension Scheme” and “Extended Paid Leave” on a five percent level, as reported in tables B.8 and B.9.[29]

Finally, Tables B.10 to Table B.15 provide a closer look on all six regressions in Table 2.4, which summarises benefits for the employer. The influence of personality tests of non-managers on team-working (Table B.10) is stable and significant across all five specifications of modern human resource practises. Even with different sets of controls, hypotheses on less monitoring (Table B.11) and reduced dismissals (Table B.12) are unrelated to personality tests of non-managers. This is also true for joint significance for manager and non-manger screening of the establishment. These negative results on dismissals across all control sets provides further evidence that personality tests are not (only) applied to improve the “fit” between applicant and firm.

Table B.13 shows significant correlations between personality tests of non-managers and “Firm Performance” for all but one control sets. The same applies to “Firm Benefit”, the compound measure whether firms benefit at least in one dimension of less monitoring, more team-working, or better performance, as depicted in Table B.14. Finally, there is a weaker relationship between screening for job candidates’ personality and “Firm Benefit 2” (Table B.15). This should not be surprising, as “Firm Benefit 2 ” is defined as “Firm Benefit” plus less dismissals. However, as shown previously, dismissals are not related to personality tests.

Summarising, Result 1 is robust to specifications 1, 4 and 5, but seems less robust regarding specifications 2 and 3, i.e. when personality tests for managers are explicitly included to the controls. However as many firms use personality tests for managers and non-managers provided that they use personality tests, both measures are highly correlated, resulting in imprecise point estimates. This is the reason why we reported adjusted Wald tests, which are by and large in line with the main regressions. Robustness tests for Result 2 do not systematically deviate from findings in the main section, suggesting that the association between personality tests and firm benefits seem to be profound.

[29]Details are available from the authors upon request.

2.4 Discussion

In previous years increasingly many contributions in personnel economics relate social preferences of employees to firm behaviour. Accounting for employees' (social) preferences may alter organisational structure within the firm and can lead to different job characteristics. (See Bartling et al. (2012a))

In this paper we use the 2004 wave of the Workplace Employment Relations Survey (WERS 2004) and find that firms behave consistent with a model gift-exchange based motivation for their employees *if* they screen job candidates for personality. We use personality tests as a proxy for the degree of reciprocity (susceptibility to gift-exchange) within the workforce. Previous research has documented that traits elicited in personality tests are correlated with (laboratory) concepts of reciprocity.

Firms which apply personality tests are more likely to provide their employees possibly non-pecuniary benefits like employer pension schemes or grant extended paid annual leave. These employers are furthermore less likely to pay very low wages and provide more on-the-job training to their employees. The topics covered in the provided training are rather general instead of workplace related, implying a higher added value for workers. Finally, there is a weak tendency that firms with personality tests are more likely to provide their employees protection against redundancies via job security. On the other hand, firms also benefit from screening for personality: we find that these firms have higher rates of team-working and are generally more successful on the market.

Importantly, competency tests upon hiring and incentive pay, both modern human resource practises similarly to personality tests, predict only poorly (if at all) benefits both for the firm as well as for employees. This implies that the use of modern human resource practises is not sufficient to explain the provision of benefits and firm performance. It is necessary that firms explicitly screen for job candidates' personality.

Closest to this study is Huang and Cappelli (2010). Using US survey data, they proxy for the importance of job candidates' "work ethic" for employers' hiring decisions. These authors find that firms which put high weight on "work ethic" on average pay higher wages, have more team-working and are more productive. Furthermore these firms monitor their employees to a lower degree and have fewer turnovers.

By and large, our analysis confirms the results of Huang and Cappelli (2010). Our results based on the WERS 2004 only deviate in two dimensions: First, we do not find stable relationships between screening and monitoring. Second, our measure of turnover – dismissals relative to firm size – is unrelated to personality tests. Note that the latter (negative) result also implies that personality tests are not primarily a device to improve the fit between applicant and firm. Together with poor predictive power of other human resource practises gift-exchange motives in firms with personality tests seem to be a plausible explanation for our findings.

Chapter 3

Good Jobs, Screening, and Labour Productivity – Evidence from the Field

3.1 Introduction

What is the optimal level of discretion employers should grant their workforce? On the one hand leaving employees a lot of leeway can increase productivity, as workers themselves might know best how to get their tasks done most efficiently.[1] This result has extensively been reviewed in Ichniowski and Shaw (2003). However, defining operating processes less strictly may induce employees to lower effort without having to fear immediate consequences. The latter argument is in line with classical moral hazard frameworks where agents maximise utility by minimising effort costs – at the expense of the principal who may for several reasons not be able or willing to fully monitor employees' actions. This basic trade-off between monitoring and shirking gave rise to fruitful research on incentive mechanisms which aim to circumvent one of the two extreme outcomes: either literally no discretion at the cost of high efficiency losses or some amount of discretion combined with high levels of shirking. To overcome that dilemma, standard theory usually suggests to pay agents for performance or, alternatively, provide efficiency wages to avoid potentially high risk premia and to cope with situations in which monitoring is particularly difficult.[2] An efficiency wage strategy is promising when both parties repeatedly interact such that the employee has a high continuation value from the relationship and the threat of dismissal is credible.

[1]The terms "worker" and "employee" are used interchangeably throughout this essay.
[2]See Prendergast (1999) for an extensive survey on incentive mechanisms.

More recently, increasingly many contributions highlight the importance of non-standard behaviour of employees. Among others, two distinct mechanisms may both lead to employees providing higher effort even in situations with moral hazard: Huang and Cappelli (2010) find that self-reported employee performance seems to be higher if managers prefer to employ job candidates with high "work ethic". It is argued that these employees unconditionally exert higher levels of effort as working hard is part of their personality. This trait makes them more desirable for firms to hire.

Great attention in recent research however was also dedicated to the concept of reciprocity which implies gift-exchange behaviour between employer and employee.[3] Employees with reciprocal traits respond to kind labour contracts – commonly achieved with higher than outside option income for them – with high levels of effort. Compared to the concept of efficiency wages, which is agnostic about (non-standard) agent behaviour and to the mechanism of "work ethic" which assumes employees to intrinsically work harder, firms attempting to make use of reciprocity need to meet two conditions: in a world with heterogeneous agents, the firm first has to employ workers with reciprocal traits and second has to offer these employees generous labour contracts. The first part can be achieved by explicit screening for reciprocity whereas the second part is most easily accomplished by paying higher than outside option incomes.

In this paper I use data from the 2004 wave of the "Workplace Employment Relations Survey" (WERS 2004), a representative survey of the British economy, to explore complementarities between human resource management practises (HRM) and labour productivity. Common sense suggests that firms deciding to grant employees high discretion may need to implement complementary practises to prevent workers from exploiting increased shirking opportunities. Contrary to this intuition I will find that firms leaving discretion do not face detrimental consequences even if discretion is not complemented with additional HRM practises which suggests that shirking may (here) not be the predominant effect. However combining high discretion with providing high income (this combination is henceforth referred to "good jobs", a terminology which is borrowed from Bartling et al. (2012a)) along with screening job candidates for personality significantly predicts higher than average labour productivity. Importantly, the necessary combination of personality tests upon hiring *and* high income suggests a conditional relationship which is consistent with gift-exchange as an underlying mechanism: generous firms may only expect high effort exertion (and hence high labour productivity) by the "right" employees, i.e. those who were screened for personality and hence should exhibit reciprocal traits. Simply providing "good jobs" alone without screening for personality is not sufficient to predict high firm performance. This study therefore casts doubts on efficiency wages as the optimal

[3] For comprehensive surveys on reciprocity, see Fehr and Gächter (2000) or Fehr and Schmidt (2006).

strategy to efficiently make use of discretion. Complementing discretion with personality tests for job candidates in hope to select the "right" employees who unconditionally exert high effort even for low wages likewise fails to be related to beneficial firm outcomes. Finally competency tests – a screening device to uncover applicant's cognitive ability – is not suitable to predict labour productivity to a similar extent as personality tests do.[4] This suggests that not screening itself but screening for personality is the key parameter to complement "good jobs", as only personality tests reveal reciprocal types.

Relating firm outcomes to generosity in income and discretion requires me to match responses from two sources of the WERS 2004 dataset. Labour productivity is taken from the management survey, which supplies firm level estimates of labour productivity. Estimates for the level of discretion and the generosity of income are derived from an employee questionnaire, for which a number of employees is drawn randomly from each firm surveyed in the WERS 2004. As income and the degree of discretion depend on several personal characteristics of the respective employee, I separately regress income and discretion on a series of observables like age, tenure, or occupational group to generate estimated values of both variables.[5] I then compare estimated and actual values of income and discretion separately to determine the degree of firm's generosity towards each single employee for each dimension. As the dataset does not provide information on individual performance or effort exertion I aggregate employee responses on firm level by calculating average scores of deviations of income and discretion for each establishment.

With personality tests and competency tests I use two distinct screening procedures job candidates may have to undergo. Though both practises are commonly used by employers, each device measures different dimensions of applicants' qualifications. Whereas competency tests are cognitive ability tests to reveal the intellectual capacity of employees, personality tests aim to uncover personal traits of the workers. Personality tests are generally based on the Five Factor Model, a theory in psychology which classifies human traits into five dimensions. These so called Big Five personality traits are "openness", "conscientiousness", "extraversion", "agreeableness", and "neuroticism". In a laboratory study, Ben-Ner et al. (2004a) relate these five traits to behaviour in dictator games and find that "agreeableness" and "openness" are positively correlated with agents' generosity in

[4] Screening usually can be distinguished in two main dimensions: screening for personality and screening for ability/competency. Interviews, computer-based tests, and reference letters are popular screening devices and are applied to both dimensions.

[5] This two-step approach of first comparing actual from estimated values and then using these deviations for further inference is similar to Black and Lynch (2001). Black and Lynch, in a first step, estimate production functions for each establishment and then relate the residual to human resource practises.

gift-exchange games.[6][7] In an earlier contribution Ashton et al. (1998) find that reciprocal altruism as measured in laboratories is closely linked to personality traits "agreeableness" and "emotional stability"; the latter refers to "neuroticism".[8] Day and Silverman (1989) furthermore find that personality traits significantly predict job performance even when cognitive ability is controlled for.

In light of this evidence, I interpret personality tests upon hiring as an indicator for the degree of reciprocity within a firm's labour force – not necessarily because employers explicitly screen for reciprocity but due to correlations between Big Five traits and laboratory measures of reciprocity. In contrast to personality tests, Ben-Ner et al. (2004a) do not find associations between ability and Big Five traits. This suggests that competency tests seem to measure traits which are orthogonal to personality traits like reciprocity.

This study at hand was inspired by a laboratory experiment by Bartling et al. (2012a). In this paper a principal sets a fixed wage and decides on the level of worker's discretion: high discretion implies high productivity of effort at the cost of unlimited shirking opportunities for the employee. Indeed the authors find that providing discretion turns out not to be profitable on its own as shirking is widely prevalent. However combining discretion with high wages renders profitable for employers (and employees) *if* employers can offer such contracts selectively to workers with high effort record. Exogenous variation in the viability of screening agents for their past effort provision allows the authors to identify screening opportunities as the causal determinant of the creation of "good jobs". If screening is permitted then this leads to the emergence of two job-clusters: "good jobs" with high wages, high discretion, and high rent-sharing and "bad jobs" with low wages, low discretion, and little rent-sharing for employees with poor reputation. Most notable, only trust and trustworthiness are necessary for the dichotomy of job-clusters to emerge endogenously.[9]

The most important difference between Bartling et al. (2012a) and this study is that I use *real-world* data. The nature of the data also implies that I cannot exogenously change the market environment as generally done by laboratory studies. Despite this limitations my results give rise to reciprocity as a plausible underlying mechanism casting doubts on

[6]"Agreeableness" refers to cooperative and compassionate behaviour towards others. "Openness" describes the degree of intellectual curiosity in comparison to cautious behaviour. For further reference, see Atkinson and Hilgard (2000).

[7]A similar finding is reported in Ben-Ner et al. (2004b). Evidence on predictability of Big Five indicators on giving can be found in Ben-Ner and Kramer (2011).

[8]See Part II of this dissertation for an extensive discussion on personality tests and reciprocity. Englmaier and Leider (2012) furthermore provide evidence for increased effectiveness of reciprocal incentives if a personality test beforehand classified agents to be reciprocal.

[9]Altmann et al. (2013) suggest a non-behavioural mechanism for labour market segregation which is the result of incomplete contracts.

efficiency wages and "work ethic" to be the sole explanations of my findings. However, this paper refrains from making causal statements: The study exploits variations which exclusively come from endogenous firms' choices of whether to make use of screening devices or not. Hence this contribution should be regarded as providing correlations between complementarities in HRM practises and labour productivity.

An asset of the WERS 2004 is that the dataset is a representative sample of the British economy, which allows making general statements. This complements the limited generalisability of laboratory experiments, in which the researcher usually does not observe behaviour of real-world decision makers. Of course, real data come at the cost that behavioural responses cannot be measured with the same accuracy in the field as in laboratory experiments: For instance, instead of using effort levels of each employee I rely on firm-wide labour productivity, which itself is a function of employees' effort provision. But regardless whether the relationship between effort and productivity is one-to-one or not, it seems plausible to regard productivity and not merely effort provision as a decisive part in managers' objective functions. Despite these potential drawbacks of real-world datasets, my results are remarkably consistent with findings of Bartling et al. (2012a).

Finally this contribution highlights the importance of personality tests in understanding interactions between the behaviour of the employer and the employee in the workplace. Altogether, my results give rise to reciprocity as a plausible explanation for the strong relationship between high firm performance, "good jobs", and screening for personality.

This paper contributes to the rich literature on workplace organisation and personnel economics. Ichniowski and Shaw (2003) provide evidence from various studies that high-performance work systems, in particular high levels of discretion, seems to turn out to be (highly) predictive for firm success whereas traditional and hierarchical workplace organisations are less correlated with favourable outcomes; in this respect this contribution is not an exception. However Ichniowski and Shaw (2003) suggest conventional methods like indoctrination of employees, management culture, or work practises which establish high effort norms and fosters peer pressure to mitigate the free-rider problem. In a study on steel finishing lines, Ichniowski et al. (1995) and Ichniowski et al. (1999) demonstrate that innovative HRM practises like incentive pay and workflows being organised in teams are associated with high line productivity.

In contrast to the above, some recent studies put more emphasis on behavioural aspects of employer-employee relationships. In a laboratory experiment Falk and Fischbacher (2006) provide evidence for workers' adverse behavioural responses after an increase in workplace control, as such policies are perceived as signals for distrust against employees ("hidden costs of control"). Even though such costs may amplify positive effects of high-performance work systems with high discretion (and are consistent with my results), this paper does not build on that mechanism. Another example for employees' non-standard reaction on institutional changes within the workplace is put forth by Nagin et al. (2002) in a field experiment on shirking in call centres. A significant fraction of employees does not react adversely to reductions in monitoring rates suggesting heterogeneity in fairness concerns among workers. These results may imply that choosing the "right" job candidates allows employers to provide discretion to their employees without facing the immediate threat of extensive shirking. This complements findings by Bartling et al. (2012b) who provide an example in which employers with fairness concerns have a broader range of contracts they are able implement as compared to selfish employers.

Similar to Bartling et al. (2012a), this study advances reciprocal responses of employees on employers' behaviour as a potential mechanism to overcome the trade-off between discretion and shirking.[10] In a theoretical contribution, Englmaier and Leider (2012) incorporate such reciprocal traits of employees into the utility function and solve an otherwise classical moral hazard problem.[11] In their framework, employing reciprocal agents gives employers the opportunity to use two distinct devices to motivate employees: Classical monetary incentives and providing the agent with higher than outside option utility, which induces effort provision through gift-exchange motives.

Closer to this study, however, are empirical papers highlighting the importance of reciprocity in the field. Leuven et al. (2005) provide survey evidence that workers with higher inclination to reciprocity have higher training rates compared to employees with low sensitivity for reciprocity. Similarly, Dohmen et al. (2009) relate individual measures for reciprocity to employee-specific labour market outcomes. They find that positive reciprocity tends to increase wages and is associated with working harder. Bellemare and Shearer (2009) report strong and persistent effects of gift-exchange for a tree-planting firm in British Columbia.[12]

Finally, this paper is related to research on screening methods of employers. Autor and Scarborough (2008) report that screening is widely prevalent among firms. This also is in line with the WERS data used in this study, where one third of all firms use personality tests and even two thirds of surveyed establishments require competency tests upon hiring. In a laboratory experiment, Englmaier et al. (2011) show that employers pay substantial wage premia for information about a worker's cognitive ability and her trustworthiness,

[10] Evidence on a relationship between effort and wage can be found in Fehr et al. (1993), Charness (2004), Cohn et al. (2012), and Kube et al. (2012).

[11] Another formal foundation of reciprocity can be found in Falk and Fischbacher (2006).

[12] Kube et al. (2013) find that employees react with negative reciprocal behaviour on wage cuts. Contrary to these studies with positive effects of reciprocity on labour market outcomes, Gneezy and List (2006) do not find strong and long lasting effects of gift-exchange in the field.

implying willingness to pay for screening devices. Finally, Wilk and Cappelli (2003) report that employers differ to a substantial extent in the level they make use of screening devices.

The remainder of the paper is organised as follows: In Section 3.2 I develop three hypotheses on labour productivity and workplace organisation. Information about the WERS 2004, the matching procedure and the estimation of income and discretion are provided in Section 3.3, along with an extensive description of the final dataset. Section 3.4 empirically tests the hypotheses with various econometric specifications. A variety of different specifications to evaluate the robustness of the results is offered in Section 3.5. Section 3.6 discusses the results.

3.2 Hypotheses

In this section I develop testable hypotheses on the relationship between HRM policies and labour productivity. First I generate job-clusters which are derived from combinations of different HRM practises and correspond to at least three major HRM policies: paying efficiency wages, screening for "work ethic" and making use of reciprocity. This classification will later on be refined to provide conditions under which reciprocity appears to be a plausible explanation.

As reviewed in Ichniowski and Shaw (2003), empirical evidence suggests that decentralisation of information flows and implementation of high-performance work systems are associated with firms being more productive. Theoretically, however, the relationship is ex-ante ambiguous as gains from decentralisation may not necessarily outweigh the increase in shirking opportunities for employees which may potentially result in overall lower effort provision. Additional empirical challenges due to the cross-sectional nature of the data arise because firms with high-performance work systems may differ in several observable and unobservable dimensions from firms with rather traditional human resource policies. In this paper I hence regard the relationship between firm performance and the implementation of high-performance work systems as an inherently empirical question which leads me to formulate the first hypothesis rather cautiously.

Hypothesis 3.1 (Discretion and Firm Performance)**.** *High-performance work systems and firm performance should be positively correlated if productivity gains outweigh potential reductions in effort provision or if firms with high-performance work systems differ in other dimensions which are positively correlated with performance.*

Even if firms with high-performance work systems may report systematically different firm performances, the focus of this paper is to identify HRM-*clusters*, which are correlated with high productivity. Inspired by Bartling et al. (2012a), I focus (next to the degree of employee discretion) on firms' generosity in wage payments and firms' use of screening methods in their hiring practises. For the latter, I distinguish between personality tests which may be correlated with agents' preferences and competency tests which are not.

If neither screening for personality nor the generosity of employees' income is correlated with labour productivity, then the data should not show any pattern in firm productivity between firms with high income provision and/or personality tests and firms without these practises. Such a pattern should be observed, when, firstly, firms are able to motivate employees independently from income levels. For instance, peer effects or effort norms could generate high effort exertion despite shirking opportunities. If secondly personality tests measure personal traits which are orthogonal to effort provision, then a firm's decision to screen job candidates should not systematically be related to performance.

Following the argument of efficiency wages, it could be sufficient to motivate workers by paying incomes which are substantially higher than obtained in comparable labour relations. This option is disregarded in static principal-agent models, as in these models interactions are one-shot. Hence this implies an abstraction from the threat of dismissing agents if their effort provision is too low. However, efficiancy wages are independent from screening devices as effort provision is not a result of agent's other-regarding preferences but is derived from inter-temporal utility maximisation.

Huang and Cappelli (2010) suggest "work ethic" to be the decisive human resource parameter to attain high firm performance. Employees with high "work ethic" unconditionally provide high effort, which makes it particularly cheap for employers to mitigate the trade-off between discretion and shirking: The employer has to pick job candidates with high "work ethic" but can provide wages that meet the agent's outside option.[13] It follows that firms that screen for personality ("work ethic") and pay low income should perform weakly better than firms paying high income, as wages are, at least in this model, unrelated to effort provision.

Finally, if employees can be motivated via gift-exchange (reciprocity), two conditions have to be met. First, an employee has to have reciprocal traits. By applying personality tests, firms either screen for reciprocity directly or screen for personal traits, which are correlated with reciprocity. The average worker hence should in these firms be more inclined to reciprocity than in firms without screening for personality. But as reciprocity is not an unconditional concept, reciprocal employees only provide high effort if the firm

[13]In fact, Huang and Cappelli (2010) find that firms screening for "work ethic" also pay higher wages. The authors explain this finding with competition among firms for these unconditionally motivated workers which finally results in rent-sharing of the generated surplus. Higher wages are hence interpreted as a consequence of high effort and not as a prerequisite for that.

previously provided a gift to them. This is achieved most easily by paying higher than outside option income. Hence, if reciprocity between employers and employees is the underlying mechanism, the data should show particularly high performance for firms which provide high wages *and* screen for personality.

These considerations result in the second hypothesis:

Hypothesis 3.2 (Job-Clusters and Firm Performance)**.** *Positive associations between high-performance work systems and firm performance should be (weakly) stronger for job-clusters in which the firm*

a) neither screens nor pays high income if employees can sufficiently be motivated otherwise,

b) pays high income if efficiency wages are sufficient to induce high effort from employees (regardless of applying screening devices),

c) screens for "work ethic" and pays low wages, if the firm can pick employees who unconditionally work hard, and

d) screens job candidates for reciprocity and pays high wages, if reciprocity is the underlying behavioural mechanism which induces employees to work hard.

Ex-ante a potential association between "good jobs", personality tests and high firm outcomes, however, is just as plausible as the explanation that screening itself, not narrowly screening for personality is sufficient to predict that pattern. As screening in general is likely to be an indicator of careful human resource policies, firms using tests upon hiring may be able to successfully provide leeway to employees and simultaneously discipline them.

However, if not screening for personality is the decisive factor for the correlation between "good jobs" and firm performance but the fact that firms use any of various possible screening devices, then the same pattern of "good jobs", screening and firm performance should arise similarly for competency tests as well. Firms using any screening devices should be aware of modern human resource practises but only personality tests are associated with traits which predict reciprocity. This insight gives rise to the third hypothesis:

Hypothesis 3.3 (Competency Tests and Firm Performance)**.** *If screening itself as opposed to narrowly screening for personality is sufficient to generate a positive association between "good jobs" and firm performance, then firms using competency tests to screen job candidates should likewise report exceptional high productivity.*

3.3 Data

The data comes from the fifth wave of the “Workplace Employment Relations Survey” (WERS 2004) with fieldwork taken place in 2004. Funded by the British government this study is part of a series with previous waves conducted in 1980, 1984, 1990, and 1998 and intents to “provide a nationally representative account of the state of employment relations and working life inside British workplaces”.[14] The data consists of four separate datasets including a management survey, an employee survey as well as a questionnaire for employee representatives and a financial performance questionnaire. The following paper studies both the management and the employee survey.

In total, the management survey consists of 2,295 Britain-based establishments from almost all branches of the economy with a minimum of five employees per firm.[15] Within each firm, a maximum of 25 employees were randomly sampled and requested to participate in the employee survey.[16] Overall, in three fourths of the establishments at least one employee returned a questionnaire – hence for 562 firms only information from the management questionnaire is available. Provided a minimum of one employee questionnaire returned, the overall response rate is slightly above 60 percent.

This paper relates firm outcomes to the following three parameters of personnel policy: The generosity of wages, the amount of discretion employees are granted, and whether the firm screens for the personality of job candidates. Whereas I retrieve information on the first two policies from the employee survey, data on personality tests and the outcome variable, labour productivity, are derived from responses by managers from the management survey. In the remainder of this section I first outline the matching procedure followed by an extensive description of the matched dataset with a special focus on the level of discretion employees enjoy as well as their income. Next, I provide information on the procedure of how wages and discretion were estimated on an individual level. I proceed by presenting details on the aggregation procedure and finally an overview of all variables of interest will be provided.

Matching A unique firm identifier allows me to match the management dataset and the employee survey, leading to a dataset which consists of 1,733 firms with 22,451 workers surveyed. The median number of workers surveyed per firm is 13 and the 25th (75th) percentile gives 8 (18) returned questionnaires per firm. Very low response rates and

[14] Source: http://www.wers2004.info/wers2004/wers2004.php. January 28, 2013.

[15] Sectors not covered by WERS 2004: Agriculture, hunting and forestry, fishing, mining and quarrying, private households with employed persons, and extra-territorial bodies.

[16] Questionnaires were distributed to all employees if the respective establishment employed less than 25 workers.

rates close to 25 are rather rare, as can be seen in Figure C.1 in Appendix C.2. The newly generated dataset is weighted with the standard weight which is provided in the employee dataset, is stratified according to the suggested procedure, and standard errors are clustered on firm level.[17]

Description of Employee-Level Variables In the employee questionnaire workers are asked to state their weekly income before taxes and other deductions by marking one out of 14 income intervals, ranging from 50 pounds a week and below to 871 pounds per week and above. The brackets are not equally spaced and the spacing increases with income.[18] Overall 414 employees refused to indicate their income and 19 multi-coded, so that 22,018 answers on income remain. A histogram of the income distribution is depicted in the left panel of Figure 3.1. The median income in the sample is 311 - 360 pounds a week (16,121 - 18,720 pounds per year). Notice however that employment relations are not necessarily full-time; about one quarter of surveyed employees work less than 30 hours a week.

To proxy for discretion I use two different measures. Employees are asked to rate whether they feel to have "a lot", "some", "a little" or no ("none") influence over what tasks they do in their job and how they do their work.[19] A histogram of both distributions is provided in the right panel of Figure 3.1. A very large fraction of surveyed employees feels to have at least some leeway over how they perform their work and which actions they can choose. However, even though only approximately 15 (25) percent state to have little or no control over the tasks they do (how they do work) a substantial difference in discretion may exist between statements of having "a lot" or only "some" leeway.

Both income and the individual level of discretion are likely to depend on several characteristics of the employee, such that absolute levels are likely to only be a poor signal about a firm's generosity – the income of an unskilled routine worker with a generous wage will in most cases still be less than that of a badly paid manager. In order to assess generosity in income and discretion, I use deviations of actual income and discretion from estimated values. However, before presenting results from the regressions I first introduce the set of control variables which are summarised in Table C.1 in Appendix C.1.

To generate indicators of generous wages and high levels of discretion, I control for an

[17] For reference, see http://www.wers2004.info/FAQ.php#5, section 5.6 "How do I apply weights and correctly estimate variances in Stata?", January 28, 2013.

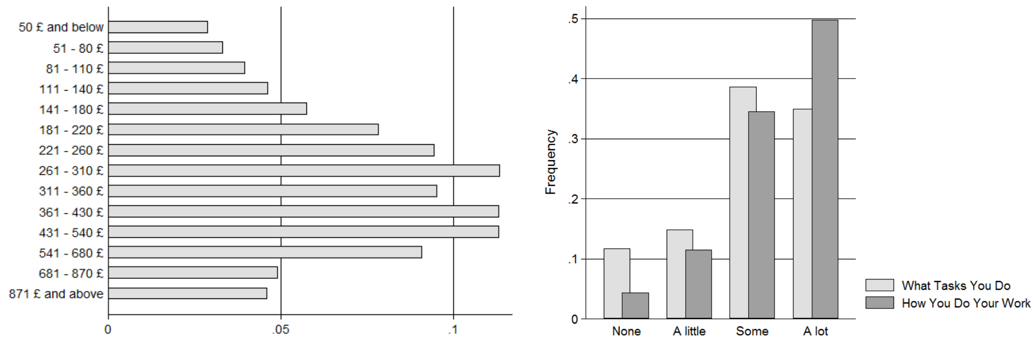
[18] The precise categories of weekly income measured in pounds are as follows: 50 and below, 51 - 80, 81 - 110, 111 - 140, 141 - 180, 181 - 220, 221 - 260, 261 - 310, 311 - 360, 361 - 430, 431 - 540, 541 - 680, 681 - 870 and 871 and above.

[19] The exact wording is "In general, how much influence do you have over the following?", followed by "What tasks you do in your job" ("What Tasks") or "How you do your work" ("How to Work").

Figure 3.1: Histograms of Income Distribution and Distribution over Perceived Discretion

This figure provides historgrams for the income distribution (measured in pounds per week) (left) and the distribution of two measures for perceived discretion (right) – the dark grey bars of panel (b) refer to preceived discretion in how employees can do their work, the light-coloured bars provide evidence on the autonomy which tasks can be done by employees. All employees who refused to answer or who multi-coded are excluded.

(a) Weekly Income (b) Discretion



employee's age and tenure, gender, academic and vocational qualifications, whether the employee is a member of a union, as well as the occupational group. Finally, I also control for the number of hours worked per week.

The median age of surveyed employees is between 40 and 49 years, with only 3.3 percent of very young (under 20) and 4.7 percent of old employees (60 and older). Employees are on average between 2 and 5 years in the respective workplace. The data furthermore distinguishes between nine occupational groups. Each employee may belong to: (1) manager and senior officials, (2) professional occupations, (3) associate and technical occupations, (4) administrative and secretarial occupations, (5) skilled trades occupations, (6) caring, leisure and other personal service occupations, (7) sales and customer service occupations, (8) process, plant and machine operatives and drivers, or (9) routine unskilled occupations. As can be seen from Table C.1, Appendix C.1.1, employees are relatively equally sampled from all occupational groups, with slightly less observations for groups (5) - (8). More than a third of the surveyed labour force is unionised and about 17 percent have been in the past. Almost half of the employees (46 percent) obtained a GCSE grade D-G (or comparable)[20], whereas only 4 percent of surveyed employees finished education with a university degree. 80 percent of employees obtained at least a level 1 NVQ (National Vocational Qualification) but less than 0.5 percent reach the top level 5, which involves substantial autonomy and includes bearing high amounts of responsibility. Gender is split almost equally with a slight concentration of males and the average working time is 36 hours per week.

[20]The General Certificate of Secondary Education (GCSE) is a necessary prerequisite for attending high school in the UK education system.

Estimation The estimation of income first has to account for censoring at the bottom and at the top and second has to take bracketing of the answers on income levels into account. For this reason I estimate a variant of a Tobit model which accounts for the survey structure of the data and allows for differently spaced intervals.[21] For the underlying (unrestricted) model of income I assume the following Mincer-type model:

$$w_i = \mathbf{S_i'}\beta_{\mathbf{S}} + \mathbf{F_i'}\beta_{\mathbf{F}} + \mathbf{X_i'}\beta_{\mathbf{X}} + \varepsilon_i. \tag{3.1}$$

where w_i is the wage employee i receives, $\mathbf{S_i}$ is a vector which contains the set of dummy variables from both academic qualification and vocational training. The vector $\mathbf{F_i}$ includes variables that describe the employee's experience, namely the tenure and the age. $\mathbf{X_i}$ finally summarises all remaining control variables I introduced beforehand, and ε_i is the error term. Tobit models imply two critical assumptions: First, it is assumed that the error is normally distributed and second a homoskedastic error structure is required, i.e. $\varepsilon_i \sim N(0, \sigma^2)$. The first issue is addressed in the robustness section by using log-income instead of income.[22] I deal with the second issue by re-estimating the wage equation with plain OLS, which does not change results.[23]

Results of the wage regression are presented in Table C.2, column (1) in Appendix C.1. Here I provide regression results for a specification that slightly deviates from the classical Mincer wage regression, by omitting the quadratic term on experience. However, both proxies for experience, age of the employee and tenure, are included as dummies for each possible category. As can be seen immediately, an inversely u-shaped relationship between age and income emerges: Ceteris paribus, employees aged between 40 and 49 have the highest income, significantly higher than the base category of under 16 - 17 year olds. With age above 49, the coefficients decrease but income stays significantly above those of job entrants. A similar picture emerges for tenure: Employees with long histories in a particular establishment have substantially higher income than employees with short tenure.

"Occupational group" has the expected influence on income: the more abstract and skill-intensive the occupation, the higher the income. The base category is managers and senior officials. Women on average earn less.

The right panel of Table C.2 provides estimates on expected discretion, which I obtain estimating the same equation 3.1 substituting income with discretion but using simple regression techniques. Column (2) gives results for the question of perceived discretion

[21] For further references see commands "intreg" and "svy" in StataCorp. (2011).
[22] Transforming income into log-income and re-estimation of the model does not change results.
[23] Regression results are available from the author upon request.

concerning what tasks employees are allowed to perform and column (3) refers to how work is done. The pattern here is again similar to what is expected: Older workers with more tenure in more abstract jobs on average report to have more leeway on what tasks to perform and also how to do them.

Aggregation and Generation of Variables Results from estimating equation 3.1 provides expected income for each employee, conditional on observables. In order to obtain an estimate of whether the employer is generous towards the respective employee with regard to income, I calculate deviations of estimated income from actual income for each employee. As however actual income is only reported within intervals, I use mean income within each interval as actual income. If an employee reports to have income in the highest or lowest category (i.e. her income is censored) I cannot calculate deviations from estimated income, because there is no sensible average for these two categories. If I, for example, set actual income of the top (open) category to its lower boundary, i.e. to 871 pounds per week, then every employee in this category with estimated income higher than 871 pounds would automatically be classified as not being paid generously, because estimated income then always exceeds actual income. In the same manner, all employees falling in the lowest category (and having estimates lower than 50 pounds) would be treated as earning higher than expected income when using the upper end of the category (i.e. 50 pounds per week). For this reason I set observations on income to missing if an employee states to be paid in the highest income category *and* the estimated income is above 871 pounds, which indicates the top (open) category; I proceed analogously with the lowest income category. For discretion I also compare actual and estimated values, but do not have to take censoring into account.

As survey responses for income and discretion are measured in brackets, this procedure is not innocuous. It implies that employees whose continuous estimate for income exactly meets the indicated interval, but exceeds the mean income of the respective interval are classified to receive higher than expected income. But as observed data do not contain any information about within category distribution, these deviations within the interval only provide tendencies towards more or less generosity in salery.

For this reason I alternatively generate a more conservative measure of generosity: Employee contracts are classified to be as expected whenever the estimated value lies in the reported interval. This implies that only large deviations of estimated and actual income lead to contracts being classified as generous. I provide further details on this procedure in Section 3.5, where I show that results of this paper do not hinge on any of the suggested specifications.

After obtaining measures for firm generosity in income and discretion for each employee, I need to aggregate this information on firm level. This is done by generating the average deviation of employees' income and discretion within each firm, such that each firm obtains a continuous score of generosity separately for income and discretion.

Subsequently I collapse the dataset by deleting all duplicate observations with regard to firms, implying that only firm level information can furthermore be used. All individual employee level information has to be aggregated on firm level at this stage. In order to additionally obtain a more compact measure than continuous firm generosity, I generate a binary variable indicating whether a firm pays higher than expected income/discretion to its employees or not. For that, I calculate the mean deviation across all firms and then relate the score of the respective firm to the average score across all firms.

Description of Firm-Level Variables Summary statistics of firm-level variables, including deviations in estimated averages of employees' income and discretion within a firm are provided in Table 3.1.

The first panel summarises the distribution of labour productivity, according to self-reported assessments of managers rating their own workplace on a five-point ordinal scale compared to competitors in the same industry.[24] As can immediately be seen, most managers regard the labour productivity of their establishment to be average or even better compared to their industry, whereas only 6.5 percent claim to perform worse. In accordance with recent literature however, this overrating bias from self-reporting seems to merely affect the absolute level but keeps the relative ordering among firms unaffected, as shown in Wall et al. (2004).[25] As the study at hand compares relative productivity and is agnostic about absolute levels, using self-reported productivity measures seems to be valid.

Human resource policy variables are summarised in the second panel of Table 3.1. The continuous measure of income generosity is centred around zero with 50 percent of the values lying between -49.5 pounds per week (i.e. less than expected) and 39.1. Reducing information to generate a binary variable of deviations in income gives a dummy variable with almost equal split.

[24] The exact question is: "Compared to other establishments in the same industry how would you assess your workplace's labour productivity?" Managers could answer the following: "A lot better than average", "better than average", "about average for the industry", "below the average", "a lot below the average" or "no comparison possible". For intuitive reasons, I re-labelled the variable, such that higher values correspond to higher productivity.

[25] Guthrie (2001) and Baer and Frese (2003) compare subjective and objective performance measures and find product-moment correlations between 0.41 and 0.81. For further evidence on self-report bias see Machin and Stewart (1996).

Table 3.1: Summary Statistics of Firm-Level Variables

	Obs.	Avg.	SD	Pctl. 25	Pctl. 50	Pctl. 75	Min.	Max.
	Outcome Variable							
Labour Productivity								
A lot better	1977	0.07	0.25	0	0	0	0	1
Better	1977	0.42	0.49	0	0	1	0	1
Average	1977	0.45	0.50	0	0	1	0	1
Worse	1977	0.06	0.24	0	0	0	0	1
A lot Worse	1977	0.004	0.06	0	0	0	0	1
	Modern Human Resource Policy							
Continuous								
Income	1728	−0.89	80.8	−49.5	−8.9	39.1	−329.8	384.7
Income: Binary	1728	0.44	0.49	0	0	1	0	1
Discretion								
What Task	1732	0.01	0.41	−0.21	0.02	0.26	−2.50	1.34
What Task: Binary	1732	0.52	0.50	0	1	1	0	1
How Work	1732	0.01	0.35	−0.18	0.04	0.22	−2.56	1.01
How Work: Binary	1732	0.54	0.50	0	1	1	0	1
Pers. Test	2292	0.34	0.47	0	0	1	0	1
Comp. Test	2291	0.61	0.49	0	1	1	0	1
	Control Variables							
Firm Size	2285	411	947.7	21	67	300	5	10006
Union	2295	0.58	0.49	0	1	1	0	1
Public Sec.	2295	0.27	0.44	0	0	1	0	1
Foreign	2295	0.02	0.14	0	0	0	0	1

Notes: This table provides information on the number of observations, mean and standard deviation, 25th, 50th and 75th percentiles as well as minimum and maximum values of firm performance, human resource practises and control variables. Statistics for each variable are calculated omitting answers "refusal", "don't know" and "not applicable", indicating unclear answers.

Both variables for discretion, i.e. the level of leeway what tasks employees do and how they do their work are centred around zero. As the distributions of both measures for discretion are highly symmetric, the binary representations of discretion have a mean close to 0.5.

About one third of the firms screen job candidates for personality, or, may (indirectly) search for reciprocal types. In contrast, over 60 percent of firms make use of competency tests aiming to elicit workers' (cognitive) ability. The correlation between both screening devices is modest: $\rho = 0.19$. About one third of the firms do not screen at all and one third only uses competency tests. Slightly below 7 percent of the firms only screen for personality and 27 percent rely on both tests. Whereas personality tests are most likely for applicants in high-skilled and abstract jobs (and sales occupations), competency tests are demanded throughout all occupational groups (see Part II of this dissertation).

The third panel refers to control variables, which aim to control for industry specific differences. More specificly, I control for firm size and firm size squared, unionisation, whether the establishment belongs to the public sector and whether it is foreignly controlled. Furthermore I include dummies for the industry and regional dummies (not part of Table 3.1).[26]

3.4 Empirical Analysis

In this section I test for the hypotheses developed in Section 3.2 with a special focus on Hypothesis 3.2, by offering three methodological approaches: First I generate HRM-clusters and relate them to firm outcomes. Second, the relationship between outcomes and HRM practises is explored in a fully-fledged regression model. Finally, this paper makes use of continuous deviations of estimated and actual values in income and discretion.

3.4.1 Discretion and Labour Productivity

Figure 3.2 provides an illustrative description of the subsequent empirical elaborations. Raw correlations provide tentative evidence for a positive association between providing high discretion to employees and labour productivity confirming previous results on positive effects of high-performance work systems on firm performance. The dashed line in Panel (a) (discretion measured via the question on leeway of "What Tasks" to perform) refers to the full dataset, i.e. including all firms and shows a slightly increasing but concave pattern. Even though I do not count these graphs as hard evidence, the correlations at least do not support a hypothesis of adverse effects of discretion on labour performance. The second measure for discretion (the level of autonomy of how to perform tasks) shows a similar pattern. A potential interpretation to these findings is that firms do not seem to excessively suffer under shirking, even in cases where employees are granted high discretion and discretion does not come along with specific human resource practises.

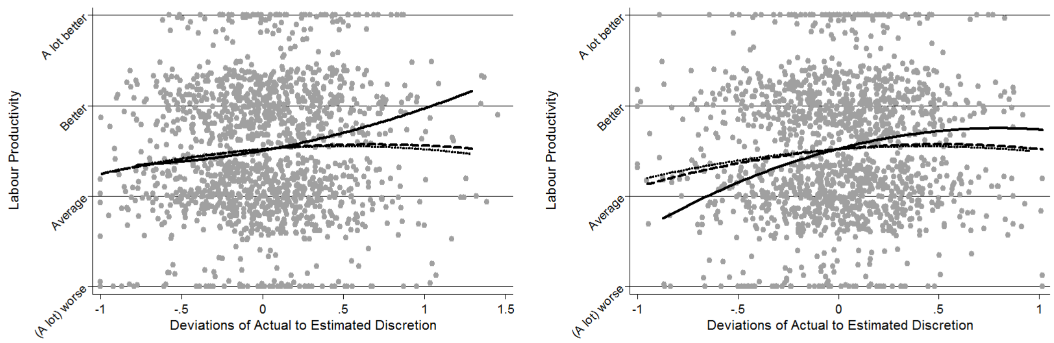
Restricting the sample to firms which pay higher than expected income and screen for personality (solid line) draws a different picture. If discretion is at most as high as expected, then firms paying high salery and screen job candidates report similar labour

[26] Industry is classified according to the UK National Statistics and distinguishes between (1) manufacturing, (2) electricity, gas, and water (3) construction, (4) wholesale and retail, (5) hotel and restaurants, (6) transport and communication, (7) financial services, (8) other business services (9) public administration, (10) education, (11) health, and (12) other community services. Regional dummies are the following: (1) North East, (2) North West, (3) Yorkshire & The Humber, (4) East Midlands, (5) West Midlands (6) East of London, (7) London, (8) South East, (9) South West, (10) Scotland and (11) Wales.

Figure 3.2: Relationship between Discretion and Firm Performance

These figures illustrate the quadratic relationship between deviations of actual to estimated levels of discretion and labour productivity. The dashed line is based on the entire sample of firms; the solid line describes the relationship for the subset of firms which pay higher than expected income and screen for personality. The dotted line uses the full sample of firms excluding firms which pay high wages and use personality tests (i.e. the subset of firms used for the solid line). For graphical reasons, I use Stata's jitter option for the scatter plot, which adds random noise to observations (the slope of the functions are unaffected).

(a) Discretion: "What Tasks" (b) Discretion: "How to Work"



productivity as the average firm (panel (a)). But whereas higher discretion only has minor effects on productivity for the average firm, establishments with high saleries and screening devices report higher labour productivity. A different pattern arises when using autonomy on how to perform tasks as measure for discretion, which is depicted in panel (b). The relationship of productivity and discretion of firms which pay high income and screen job candidates is similarly concave but has a steeper slope than the average firm. This implies that if employees in high-income and screening establishments do not receive high levels of discretion, labour productivity sharply decreases. These establishments are successful in terms of labour productivity only if they grant high discretion to their employees.

The dotted line corresponds to the full sample excluding firms which pay high income and screen for personality. I provide these estimates to show that the slight positive slope of the full sample (dashed line) is not driven by the subset of firms, which pay high income and screen for personality (solid line).

Result 3.1 (Discretion and Firm Performance)**.** *Firms have slightly higher labour productivity if their employees enjoy high levels of discretion.*

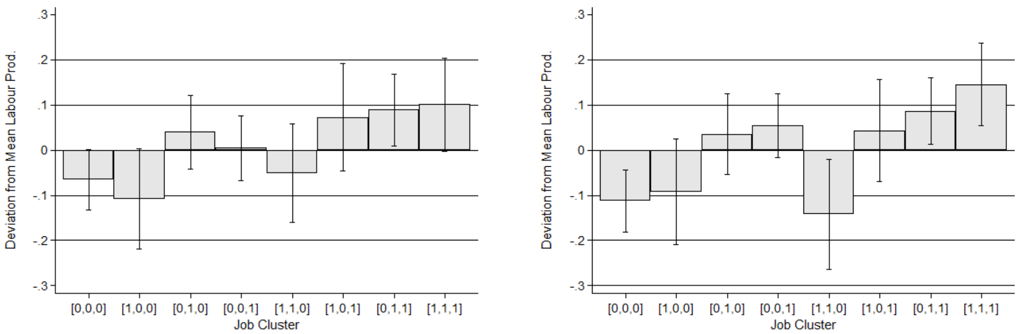
3.4.2 Interacting Income, Discretion, and Personality Tests

HRM-Clusters – Graphical Approach In this section I use the binary representation for income and discretion, as explained in Section 3.3, where each establishment is

Figure 3.3: Firm Performance and Human Resource Cluster

Panel (a) and (b) provide deviations and 90 percent confidence bands of mean firm performance for firms classified in 8 different HRM-clusters. Read classification as follow: [personality tests, high income, high discretion] i.e. the first position displays whether firms in this cluster use personality tests upon hiring (1) or not (0). The second position refers to whether firms in this cluster pay higher than average expected income (1) or not (0) and the last entry is related to firms with higher (1) or lower (0) than mean expected discretion.

(a) Discretion: "What Tasks" (b) Discretion: "How to Work"



classified to either pay high or low income and to grant much or little discretion, respectively. I combine this information with firms' screening methods for personality (likewise binary) to assign every single firm to one of eight HRM-clusters. These clusters are all potential combinations of three binary variables implying $2 \times 2 \times 2$ combinations, ranging from firms with (on average) lower than expected wages, lower than expected discretion, and no screening for personality to firms with high wages, high levels of discretion, and personality tests for job candidates.

Figure 3.3 provides deviations from mean firm performance for each HRM-cluster to the overall performance average of all firms in the dataset.[27] Cluster on the x-axis refer to the following notation: [personality tests, high income, high discretion], where "1" in the respective position implies that firms requires personality tests upon hiring, pay high income or allow for discretion. To the extremes, the illustration shows firms which do not screen for personality, pay low income and do not grant high discretion on the very left of each panel ([0,0,0]) and firms with personality tests, high income and discretion to the right end ([1,1,1]).

As can immediately be seen in both panels, firms in cluster [1,1,1], i.e. establishments providing "good jobs" and screen for personality, report the highest labour productivity which lies significantly above average productivity. In accordance to Result 3.1, clusters of both panels are split into two groups: All clusters, in which firms provide higher than estimated discretion perform above average (despite not necessarily significantly higher),

[27]Relative frequencies of each of the eight HRM-clusters are provided in Figure C.2 in the Appendix, Section C.2.

whereas firms in 3 out of 4 clusters in both panels with little discretion report to have poorer than average labour productivity. Firms using personality tests and pay high income but do not allow their employees high levels of discretion (cluster [1,1,0]) report exceptionally low labour productivity in panel (b).

HRM-Cluster – Regression Approach Equation 3.2 provides more structure than simple mean comparisons. I estimate for each firm j the influence of belonging to a certain HRM-cluster, $\mathbf{G_j}$, on ordered outcome variable y_j. The vector $\mathbf{G_j}$ includes seven dummy variables for all but one potential HRM-cluster. $\mathbf{X_j}$ contains control variables, including size of the establishment (number of employees), size squared and dummies for whether the firm is unionised, is owned by a foreign company and belongs to the public sector. Finally, the vector includes dummies for industry and region as described in footnote 26. More precisely, I estimate the following reduced form model:

$$y_j = \mathbf{G_j'}\beta_{\mathbf{G}} + \mathbf{X_j'}\beta_{\mathbf{X}} + \varepsilon_j. \tag{3.2}$$

Simple regression analysis of equation 3.2 confirms results from previous mean comparisons.[28] Using firms without personality tests, low income and limited discretion, i.e. cluster [0,0,0] as base category (columns (1) and (3)), the job-cluster which refers to "good jobs" and screening ([1,1,1]) predicts significantly higher labour productivity for both measures of discretion. In accordance to panel (b) in Figure 3.3, all three remaining job-clusters with high discretion concerning employees' autonomy on how to carry out their work ([0,0,1], [1,0,1] and [0,1,1]) yield significantly higher labour productivity as compared to cluster [0,0,0] (column (3)).

Columns (2) and (4) provide results for the same regression but relative to cluster [1,1,1] as baseline category. It is important to notice that all job-clusters yield significantly worse labour productivity, when discretion refers to which tasks employees are allowed to do (column (2)). This implies that also within the subset of rather successful firms using high discretion (Result 3.1), paying high wages and *simultaneously* screening for personality is associated with significantly higher labour productivity.

[28]Throughout this paper I use simple regression techniques, because an ordered probit approach yields qualitatively the same results. As simple regression analysis facilitates the interpretation, I decided to report these results. Ordered probit results are available from the author on request.

Table 3.2: Regression of Labour Productivity on HRM-Cluster

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
HRM Cluster				
[0,0,0]		−0.50***		−0.50***
		(0.14)		(0.16)
[1,0,0]	0.100	−0.40**	0.13	−0.36*
	(0.15)	(0.18)	(0.15)	(0.20)
[0,1,0]	0.21**	−0.29**	0.29***	−0.20
	(0.10)	(0.14)	(0.10)	(0.17)
[0,0,1]	0.13	−0.37***	0.26***	−0.24
	(0.10)	(0.14)	(0.093)	(0.16)
[1,1,0]	−0.065	−0.57**	0.019	−0.48*
	(0.25)	(0.26)	(0.25)	(0.28)
[1,0,1]	0.21	−0.29*	0.32**	−0.17
	(0.14)	(0.17)	(0.13)	(0.19)
[0,1,1]	0.13	−0.37***	0.16*	−0.33**
	(0.11)	(0.14)	(0.098)	(0.16)
[1,1,1]	0.50***		0.50***	
	(0.14)		(0.16)	
Union	−0.064	−0.064	−0.042	−0.042
	(0.084)	(0.084)	(0.084)	(0.084)
Pub. Sector	−0.068	−0.068	−0.060	−0.060
	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.55***	0.55***	0.64***	0.64***
	(0.12)	(0.12)	(0.13)	(0.13)
Constant	3.61***	4.11***	3.52***	4.02***
	(0.20)	(0.22)	(0.20)	(0.24)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1812	1812	1812	1812
R^2	0.093	0.093	0.103	0.103

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work". Columns (1) and (3) use cluster [0,0,0] as base category; column (2) and (4) omit cluster [1,1,1].
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This result similarly applies to the second measure of discretion. Although a few job-clusters are not significantly distinguishable from job-clusters of "good jobs" and personality tests, point estimates are exclusively negative.

Fully-Fledged Interaction Model In the following, I depart from the rather inflexible analysis of eight pre-specified job-clusters and allow for a fully-fledged interaction model of personality tests, income and discretion. To facilitate interpretation the following analysis is based on binary HRM measures. Subsequently I allow HRM practises to be measured continuously.

The following model measures the relationship of high income ($I_j = 1$), high discretion ($D_j = 1$), and compulsory personality tests upon hiring ($P_j = 1$) on firm performance y_j for each firm j. Apart from the main effects of HRM practises on performance, interaction terms are of main interest as these effects provide insight into complementarities of HRM practises: $I_j \times D_j$ is defined as the interaction between income and discretion and yields value 1 if firm j both provides high income and leaves discretion to their employees (provide "good jobs"). $I_j \times P_j$, the interaction between high income and compulsory personality tests and $D_j \times P_j$, which interacts high discretion with personality tests are defined accordingly. The three-way interaction $I_j \times D_j \times P_j$ is equal to 1 for all firms which were previously classified to provide "good jobs" *and* additionally demand personality tests when recruiting new employees. Finally, the vector $\mathbf{X_j}$ contains control variables as defined previously in this section.

$$y_j = \beta + \beta_I I_j + \beta_D D_j + \beta_P P_j + \beta_{ID}(I_j \times D_j) + \beta_{IP}(I_j \times P_j) + \beta_{DP}(D_j \times P_j) \\ + \beta_{IDP}(I_j \times D_j \times P_j) + \mathbf{X_j}'\beta_{\mathbf{X}} + \varepsilon_j \quad (3.3)$$

Table 3.3 provides estimation results of model 3.3 using simple regression analysis.[29] As all HRM practises are defined on a binary support, the constant can intuitively be interpreted as the labour productivity of the average firm (average with respect to controls) which neither pays high income nor provides discretion nor screens for personality. The main effect of income is the difference in reported labour productivity of firms which do not pay high income compared to establishments which do pay high income (everything else equal). If the ceteris paribus condition implies that firms neither grant high levels of discretion nor screen for personality then the coefficient on income displays the full effect of high income on productivity. The positive and significant coefficient on income for all specifications hence provides evidence that firms paying high income ceteris paribus report higher levels of labour productivity as compared to firms which pay low income. Similarly a positive and significant coefficient on discretion in 3 out of 4 specifications reflects a positive correlation between discretion and labour productivity. Only the main effect of personality tests is not significantly different from zero.[30]

[29]Here, again, applying ordered probit estimation qualitatively does not change the results.

[30]In Part II of this dissertation, Englmaier, Kolaska, and Leider in fact find a positive relationship

Table 3.3: Regressions of Labour Productivity on HRM Complementarities

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
High Income (I)	0.24**	0.21**	0.29***	0.29***
	(0.11)	(0.10)	(0.11)	(0.10)
High Discretion (D)	0.18*	0.13	0.31***	0.26***
	(0.10)	(0.10)	(0.099)	(0.093)
Pers. Test (P)	0.12	0.100	0.15	0.13
	(0.15)	(0.15)	(0.15)	(0.15)
I × D	−0.24	−0.22	−0.37**	−0.39***
	(0.15)	(0.15)	(0.15)	(0.15)
I × P	−0.37	−0.37	−0.40	−0.41
	(0.29)	(0.29)	(0.30)	(0.29)
D × P	−0.041	−0.024	−0.089	−0.066
	(0.21)	(0.20)	(0.21)	(0.20)
I × D × P	0.71**	0.67*	0.72*	0.67*
	(0.36)	(0.35)	(0.38)	(0.36)
Union		−0.064		−0.042
		(0.084)		(0.084)
Pub. Sector		−0.068		−0.060
		(0.11)		(0.11)
Foreign		0.55***		0.64***
		(0.12)		(0.13)
Constant	3.39***	3.61***	3.34***	3.52***
	(0.073)	(0.20)	(0.069)	(0.20)
Firm Controls	*No*	*Yes*	*No*	*Yes*
Subpop. Observations	1815	1812	1815	1812
R^2	0.023	0.093	0.037	0.103

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests and its interactions. The first panel (column (1) and (2)) refers to answers on the question “What Tasks” as proxy for discretion, columns (3) and (4) use “How to Work”.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

The two-way interactions $I_j \times D_j$ is significantly negative for discretion measured in employees' leeway how to work. This means that firms paying high income and leave their workers discretion but do not screen for personality report significantly lower productivity of labour than firms with limited discretion all other parameters equal. Applying a causal interpretation, $I_j \times D_j$ is the additional effect from complementarities between income and discretion when switching from low to high wages in firms which already provide

between firm performance and personality tests. As model specifications are very different, results are not directly comparable.

high levels of discretion. As the interaction is significantly negative, this policy, which corresponds to paying efficiency wages, turns out to be not successful. The interaction of high income and personality tests (paying low income) $D_j \times P_j$ is not significantly distinguishable from zero, discarding optimality of the strategy to rely solely on "work ethic" but refrain from paying high wages. Finally the third interaction, $I_j \times P_j$ does similarly not depict a significant pattern for any measure of discretion.

The main interest however, is placed on the three-way interaction term between income, discretion and personality tests, $I_j \times D_j \times P_j$. This interaction is significant in all four specifications. Intuitively (and causally interpreted) this interaction is the additional effect of introducing personality tests given that the respective firm already pays high income and leaves their employees high discretion on productivity. The overall effect of personality tests (given the firm offers "good jobs") is the sum of all coefficients, which contain personality tests, i.e. $\beta_P + \beta_{IP} + \beta_{DP} + \beta_{IDP}$. A Wald test on the null hypothesis that the sum of all four coefficients containing personality tests is zero can be rejected in three out of four specifications.[31]

Analysis with Continuous Deviations A potential immediate critique of the previous analysis is that it entirely relies on the binary classification of firms' HRM practises, which implies that a lot of information is (unnecessarily) lost. In the following, I exploit continuous deviations of actual income and discretion to estimated outcomes. However, this procedure also comes at a cost. As answers on the survey question concerning discretion are ordinal, deviations are not directly interpretable, suggesting a very cautious interpretation of the subsequent results.

In order to address complementarities of different HRM practises, it is essential to include interactions, as done before. However as multiple interaction effects of continuous variables are very problematic to interpret, I keep the analysis tractable by dividing the dataset into firms using personality tests upon hiring ($P_j = 1$) and those who do not ($P_j = 0$). For each dataset $k \in \{0, 1\}$, I then separately estimate the effect of income i_j, discretion d_j and the multiplication term ($i_j \times d_j$) of income and discretion on firm performance y_j.

[31] The models which include firm controls (columns (2) and (4)) strongly reject the null at a 1%-significance level and at a 5%-significance level, respectively. Specification (1) rejects the null at a 10%-level and model (3) fails to reject it (Prob > F = 0.18).

Table 3.4: Regression of Labour Productivity on HRM Complementarities II

	What Tasks		How Work	
	w/o PT (1)	w/ PT (2)	w/o PT (3)	w/ PT (4)
Income	−0.000086	0.00066	−0.00022	0.00070
	(0.00059)	(0.00075)	(0.00057)	(0.00074)
Discretion	0.026	0.30*	0.064	0.48**
	(0.094)	(0.17)	(0.089)	(0.24)
Income × Discretion	−0.00094	0.0019	−0.0011	0.0044**
	(0.00065)	(0.0015)	(0.00073)	(0.0019)
Union	−0.042	−0.13	−0.024	−0.13
	(0.095)	(0.18)	(0.097)	(0.17)
Pub. Sector	−0.095	0.052	−0.093	0.015
	(0.12)	(0.20)	(0.12)	(0.20)
Foreign	0.48***	0.88***	0.50***	0.59**
	(0.14)	(0.24)	(0.14)	(0.24)
Constant	3.75***	3.71***	3.73***	3.70***
	(0.22)	(0.27)	(0.22)	(0.28)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	998	496	998	496
R^2	0.088	0.249	0.092	0.266

Notes: This table provides linear regression coefficients and standard errors of labour productivity on income, discretion and the interaction between income and discretion. The first panel (column (1) and (2)) uses answers on the question “What Tasks” as a proxy for discretion, columns (3) and (4) use “How to Work”. Columns (1) and (3) correspond to the restricted set of firms which do not use personality tests, columns (2) and (4) refer to establishments which make use of personality tests for job candidates. Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

I estimate the following reduced-form model:

$$y_{j|P_j=k} = \beta_i i_j + \beta_d d_j + \beta_{id}(i_j \times d_j) + \mathbf{X_j'}\beta_{\mathbf{X}} + \varepsilon_j \qquad \forall\, k \in \{0,1\}. \tag{3.4}$$

Equation 3.4 is estimated with simple regression techniques.[32] As both income and discretion are centred around zero, the intercept of each regression depicts mean labour productivity if actual and estimated values of income and discretion are exactly aligned (meaning that firms exactly provide wages and discretion as expected), at the mean of all control variables. β_d can be interpreted as the average difference in reported labour productivity of the mean firm (with regard to controls) if employees on average report one step higher perceived discretion and income is exactly as expected (i.e. $i_j = 0$, leading

[32]Here, again, applying ordered probit estimation does not alter the results.

the interaction term to be zero). The same is true for $i_j > 0$ and $d_j = 0$; the main effect of income can be seen as the impact of paying one pound more per week on average to each employee on labour productivity, given that the firms allow discretion exactly as expected. The interaction term is hence only different from zero, if both HRM measures depart from expected values. A positive interaction can be interpreted as follows: The more actual income exceeds estimated income, the stronger the association of discretion on firm performance becomes. The interpretation vice versa is true as well, implying that the more "excess" (i.e. above estimated) discretion firms allow their employees on average, the more increasing income is associated with firm outcomes. As in a framework with continuous wage and measures for discretion no firm exhibits income and discretion exactly at the sample mean, the effect of income on firm output depends on the level of discretion if the interaction effect is significantly different from zero.

Columns (1) and (3) of Table 3.4 provide evidence on estimates for income, discretion and the interaction of both for the subset of firms, which do not use personality tests when screening job candidates. For this subset of firms, neither income nor discretion (or the interaction) is associated with labour productivity. This is true for control variables as well with the exception of firms which are foreign owned: These firms report significantly higher labour productivity.

Restricting the dataset to firms which screen for personality reduces the number of observations, as can be seen in column (2) and column (4). Whereas the main coefficients of income increase but are indistinguishable from zero even if firms screen for personality, coefficients of discretion sharply increase 10-fold and 7.5-fold. Likewise point estimates of the interaction term increase compared to firms without personality tests and result in significant correlation of the interaction between discretion and income, and labour productivity in column (4). Using leeway on what task to perform as measure for discretion and conditioning on firms with personality tests also increases point estimates of the interaction term, as can be seen in column (2). Higher standard errors however, lead to insignificant coefficients, which may be a result of limited sample size. It is important to notice that also the goodness-of-fit is approximately three times as high when conditioning the sample on firms with personality tests but otherwise estimating the identical model. This provides evidence for high explanatory power of HRM practises for labour productivity.

To show that the impact of discretion on labour productivity crucially depends on the level of income if firms screen for personality, I finally provide estimates for the evolution of discretion along different levels of income deviations. Estimates in Table 3.5 are simply the linear combination of the main effect of discretion on labour productivity for three

Table 3.5: Impact of Discretion on Labour Productivity

	Coeff.	Std. Err.	t	p>|t|	95% Conf. Int.	
With Personality Test						
What Tasks						
Slope at Mean of Income	0.30	0.17	1.78	0.08	−0.03	0.64
Slope at Mean + 1 SD of Income	0.46	0.23	2.01	0.05	0.01	0.91
Slope at Mean − 1 SD of Income	0.15	0.18	0.80	0.42	−0.21	0.50
How Work						
Slope at Mean of Income	0.48	0.24	1.98	0.05	0.005	0.95
Slope at Mean + 1 SD of Income	0.83	0.37	2.26	0.02	0.11	1.55
Slope at Mean − 1 SD of Income	0.12	0.17	0.74	0.46	−0.20	0.45
Without Personality Test						
What Tasks						
Slope at Mean of Income	0.03	0.09	0.28	0.78	−0.16	0.21
Slope at Mean + 1 SD of Income	−0.05	0.12	−0.41	0.69	−0.29	0.19
Slope at Mean − 1 SD of Income	0.10	0.09	1.12	0.26	−0.08	0.28
How Work						
Slope at Mean of Income	0.06	0.09	0.73	0.47	−0.11	0.24
Slope at Mean + 1 SD of Income	−0.03	0.12	−0.21	0.83	−0.27	0.22
Slope at Mean − 1 SD of Income	0.15	0.09	1.81	0.07	−0.01	0.32

Notes: This table provides linear regression coefficients, standard errors and confidence intervals for the effect of discretion on labour productivity for different levels of income. The first line of each panel is the compound effect (main effect plus interaction) of discretion at mean income payment. The second line describes the effect of discretion on productivity if income is one standard deviation above the mean. The third line corresponds to estimates for income one standard deviation below the mean. The first panel reports estimates for firms using personality tests, the second panel for firms without screening for personality.

different levels of income: income evaluated at its mean (i.e. no deviation from estimated income), actual income being one standard deviation above and one standard deviation below estimation. The first panel provides estimates for the subset of firms, which screen for personality and the second panel for firms not using compulsory personality tests.

The effect of discretion on labour productivity is largest when income is high and drops considerably with low income for firms which screen job candidates for personality. This is true for both measures of discretion. The coefficient for discretion, given that income is one standard deviation below its mean, is not distinguishable from zero; in that case an increase in discretion does not predict higher labour productivity. As before, the subset of firms without personality tests do not show a clear pattern. I interpret this as personality tests being crucial for "good jobs" to translate into high labour productivity.

Result 3.2 (Job-Clusters and Firm Performance)**.** *Firms paying higher than expected income to their employees while at the same time allowing their workforce substantial discretion on how to perform tasks (HRM practises which are associated with "good jobs") report significantly higher labour productivity. However, this is the case only if these firms screen job candidates for personality. Hence, this pattern is consistent with reciprocity as underlying mechanism to achieve high firm performance, discarding explanation (a) no additional human resource practise necessary, (b) efficiency wages only and (c) relying on "work ethic" as sole explanations.*

3.4.3 Competency Tests

Firms offering "good jobs" report to have significantly higher labour productivity if they screen for personality when hiring employees. However, personality tests may only be a proxy for advanced HRM practises which itself is likely to be positively correlated with firm performance. Competency tests may similarly proxy for firms with advanced HRM policies but ability tests unlike personality tests aim not to uncover personal traits of the employee but try to reveal her ability. Though both tests are indicators of modern HRM practises of a firm, only personality tests are consistent with reciprocity as an enhancing mechanism for productivity. Hence, this section summarises the results using the same analysis as in the previous section but defining competency tests instead of personality tests as a screening device.

Table C.3 (Appendix, Section C.1.2) provides estimates from equation 3.2, where personality tests P_j are replaced by competency tests C_j. Column (1) and (3) report significantly higher labour productivity if firms provide "good jobs" and screen for ability compared to firms in cluster [0,0,0]. Furthermore using firms with "good jobs" and ability tests as base category I find significantly higher levels of labour productivity compared to establishments in other job-clusters. This effect is particularly strong for discretion measured by the question on how autonomous employees are allowed to perform their tasks. Even though results in these regressions are considerably weaker than results on personality tests, using this piece of evidence does not speak against an association of labour productivity and "good jobs" combined with competency tests.

Results on estimating three-way interactions of model 3.3 are provided in Table C.4. Discretion relates positively to labour productivity and the coefficient is substantial if discretion is measured by employees' leeway on how to work. More importantly however, the interaction effect of high income, high discretion, and competency tests is indistinguishable from zero for all four specifications. Intuitively this means that firms providing "good jobs" and screen for competency when hiring employees do not report higher labour productivity than firms offering equally good jobs but do not screen job candidates for ability. This is in sharp contrast to findings on the impact of personality tests which renders screening itself unlikely to be the driving force of "good jobs" to translate into high labour productivity.

Result 3.3 (Competency Tests and Firm Performance). *Screening job candidates for competency as opposed to personality has considerably less power to explain labour performance if firms offer "good jobs". Not screening itself but screening for personality is decisive for "good jobs" to turn into high labour productivity.*

3.5 Robustness

In this section I perform a series of robustness checks: First, instead of estimating deviations in income and discretion, I use raw responses on income and discretion. Secondly, I only classify firms to be generous if actual income/discretion is at least one category higher than estimated values. In a third check, I perform the Tobit model substituting income with logarithmised income. Fourth, in order to provide a different method of how to aggregate employee responses into firm averages, I first calculate binary firm generosity towards each employee and second aggregate over all employees per establishment. To address the fact that some firms simultaneously make use of personality and ability tests, I finally re-estimate models on both screening devices excluding firms which simultaneously search for personality and ability.

Raw Responses In this section, I re-estimate models 3.2 and 3.3 using raw responses from the employee questionnaire. By skipping the entire procedure on estimating income and discretion I hence do not account for personal characteristics of each employee. However, because firms differ in the composition of *which* employees answer the questionnaire I may systematically over- and underrate income and discretion of firms. If questionnaires however were returned purely random in each establishment, then using raw correlations should only increase noise but is not expected to systematically bias coefficients.

As shown in Table C.5, column (1) and (2) in the Appendix, Section C.1.3, estimates for both measures of discretion show a similar pattern as in the main section, even though coefficients are estimated less precisely. For model 3.2 and using job-clusters of firms providing "good jobs" and personality tests as reference category we find consistently negative coefficients; some clusters exhibit significant lower associations with labour productivity. Column (1) and (2) in Table C.7 finally re-estimates model 3.3 substituting estimated income and discretion with raw values. Here, we find significant interaction terms between "good jobs" and personality tests only for one of two specifications.

Categorising Deviations Another potential critique could be the use of (hypothetical) continuous deviations between estimated and actual values for income and discretion. For the previous analysis I generate a continuous measure of expected income, though actual income is only observed within intervals. This, however, implies that firms are classified to pay less generous income to a certain employee if the estimated income is higher than the average income within the respective interval. Some employees hence will falsely be categorised to receive generous wages. The situation is analogously when estimated income is below the mean income within an interval, which leads firms to appear generous. Despite inaccuracies, however, the procedure so far is not expected to systematically bias the results (assuming a symmetric income distribution around the mean of each category) but having rather a tendency to increase the variance.

To nevertheless address these concerns I subsequently classify firms only then to be generous with regard to income if the lower bound of their actual income interval exceeds their estimated income. All employees, whose estimated income lies in the interval of their actual payments are classified as to receive income as expected. If firms pay as expected then deviations are equalised to be zero; if establishments pay higher (lower) than expected then I use the mean payment of the actual interval as reference payment. I similarly proceed with estimates on discretion, where employees' answers on perceived discretion are scales from "1" ("none") to "4" ("a lot").[33] Given this scale, I only count answers to be higher (lower) than expected if expected and actual values deviate by at least the positive (negative) magnitude of 0.5. In these cases actual discretion is not the closest integer to estimated discretion.

Table C.5 (column (3) and (4)) provides estimates from model 3.2 for both suggested measures for discretion. Contrary to the main results, in column (3) the cluster [1,1,1] is not significantly stronger correlated with productivity compared to some other cluster, though all point estimates direct towards this relationship. Using responses on the question "How to Work" as measure for discretion, estimates are comparable to the main results (column (4)). A similar picture arises for model 3.3 in Table C.7. Human resource indicators do not show significant patterns (column (3)) using "What Tasks" as measure for discretion – the magnitude of the coefficient on "good jobs" in combination with screening for personality is considerable smaller than in the base specification. Applying "How to Work" as measure for discretion, results are again comparable to the main section.

[33]Issues on the interpretability of ordinal responses are dealt with in "Alternative Method of Aggregation" in this section.

Log-Income Crucial for the estimation of income with interval regressions is the normality assumption. As the distribution of log-income in many applications is closer to the Gaussian normal (compared to raw income) I re-estimate model 3.1 after transforming income into log-income.

The transformation has no effect on the results of job-clusters (Table C.5 in Appendix C.1.3): Relating labour productivity to HRM-clusters I find negative point estimates for all job-clusters when omitting firms with "good jobs" and personality tests as base category. Six out of seven clusters (for each measure of discretion) report significantly lower productivity. Results on model 3.3 are presented in Table C.7. Here, however, I do not find significant interaction effects on $I_j \times D_j \times P_j$ for discretion measured by the question on "What Tasks". Using "How to Work", I find high and significant correlations between the three-way interaction and labour productivity. Hence, the findings in this section affirm that results from the main section are robust to a transformation of income into log-income.

Alternative Method of Aggregation A serious concern of the previous analysis is the interpretation of deviations from estimated HRM practises, although discretion is an ordinal measure without scale. Moreover, when calculating firm averages, exceptionally high positive (negative) deviations of discretion for one employee could potentially offset lower (higher) than expected levels for several employees.

Here, I address both problems by changing the order of aggregation and averaging across firms. Before, I summed deviations in income (discretion) across all employees within one establishment and then took the mean deviation in income (discretion) within the firm. In this section I first calculate a binary measure for each single employee, indicating whether this respective employee receives generous income (discretion) or not from her firm. In a second step I calculate the fraction of employees with high income (discretion) for each firm. Finally, a firm is classified to be generous with regard to income (discretion) if it provides more employees high income (discretion) than the average firm in the sample.

Estimates qualitatively do not change when applying the latter method of aggregation. Columns (1) and (2) in Table C.6 provide estimates for model 3.2 using job-cluster [1,1,1] as base category. As seen immediately almost all job-clusters yield significant negative correlations with labour productivity compared to firms which provide "good jobs" and screen for personality. In Table C.8 (columns (1) and (2)) I find positive and significant correlations of productivity with the interaction of "good jobs" and screening, $I_j \times D_j \times P_j$. This implies that the effect of HRM practises on productivity is not driven by the way how I aggregate information from the regressions on income and discretion onto firm level.

Excluding Firms with Personality and Competency Tests A final potential worry of the previous analysis is the fact that some firms use personality tests and competency tests simultaneously. In this paragraph I exclude exactly these firms from the analysis. This may be particularly interesting for the results on competency tests because these results include a subset of firms which also screen for personality.

Estimation results on job-clusters excluding the subset of firms using both screening devices are provided in Table C.6, where columns (3) and (4) refer to personality tests as a screening device and columns (5) and (6) to ability tests. Comparing results to the main tables, I do not find qualitative differences. This is only partly true for model 3.3. As can be seen in columns (3) and (4) the interactions of "good jobs" and personality tests are not significantly different from zero for both measures of discretion, even though point estimates sharply increase. This observation however, may be explained by limited sample size resulting in standard errors which are approximately 50 percent higher compared to the base specification. Finally, competency tests (columns (5) and (6)) in combination with "good jobs" cannot explain labour productivity, which is in line with Result 3.3. Here, standard errors are comparable to the standard errors in the main regressions.

3.6 Discussion

Research in personnel economics has highlighted the importance of workplace organisation for firm success. Rather recently, however, a number of studies find that behavioural aspects within firms may shape outcomes. This implies that taking the "right" actions may allow employers to benefit from non-standard behaviour of employees. Three of these potentially "right" actions are presented in this paper.

This paper uses field evidence from the "Workplace Employment Relations Survey" to relate three human resource policies – paying high income to employees, leaving worker high discretion, and screening for personality or competency – to firm performance. I show that firms which pay high income, grant high level of discretion, and screen their job candidates for personal traits report to have exceptional high labour productivity. I interpret this finding as evidence consistent with employees responding to gift-exchange; job-clusters which are associated with efficiency wages or high "work ethic" of employees alone are not associated with positive firm outcomes. This similarly applies for firms which screen for ability instead of personality.

In a broader context, this analysis shows the importance of personality tests when screening job candidates. Interestingly, however, in this dataset only one third of the firms make use of that screening device, which is a bit of a puzzle: If personality tests are the key to increase labour productivity (because this device reduces the number of employees with adverse behaviour towards the firm) then one should expect firms to increasingly make use of screening for personality. If however, only a limited number of employees in the population exhibit reciprocal traits then rising demand for these workers could lead to segmentation in the labour market: Successful firms with reciprocal employees and "good jobs" on one side and firms providing jobs with low payments and low discretion on the other. (The argument of segmentation in labour markets has, in a slightly different context, already been made by Bartling et al. (2012a))

A natural next step could be to provide causal evidence of HRM practises, in particular of personality tests on firm performance using field data. Whereas laboratory studies can isolate underlying principles, it is often not clear whether the identified mechanism has real-world implications. Gaining evidence on the actual importance of reciprocal behaviour between employer and employee could improve labour market relations with potential benefits for both parties.

Appendix A

Appendix to Chapter 1

A.1 Proofs

A.1.1 Proof of Lemma 1.1

We begin the proof by making two observations. First, it is easy to see that $F^R(f_t, \eta)$ is strictly decreasing and convex in η due to strict concavity of $u(\cdot)$. Hence, the fixed point $\tilde{\eta}$ satisfying $F^R(f_t, \tilde{\eta}) = 0$, which exists by assumption 1.1 (i) is unique. Second, note that for all $\eta > w_t$ we have that

$$F^P(f_t, w_t, \eta) < (1 - H(\eta)) \Big((1-\alpha)\, E[u(w_\tau)|\, w_\tau \geq \eta, f_t] + \alpha u(\eta) - u(\eta)\Big) - c \leq F^R(f_t, \eta),$$

since $E[u(w_\tau)|\, w_\tau \geq \eta, f_t] \geq u(\eta)$. Thus, for large η, $F^R(\cdot)$ is an upper bound for $F^P(\cdot)$, where the former is smaller than zero for $\eta > \tilde{\eta}$.

Now, note that there always exists a fixed point $\bar{\eta}$ for which $F^P(f_t, w_t, \bar{\eta}) = 0$. To see this, consider the cases $\alpha = 0$ and $\alpha = 1$. For $\alpha = 0$, $F^P(f_t, w_t, \eta_t) = F^R(f_t, \eta_t)$ and by assumption 1.1 (i) there always exists η' satisfying $F^R(f_t, \eta') = 0$. For $\alpha = 1$, $F^P(f_t, w_t, \eta) = (1 - H(\eta))\, (u(w_t) - u(\eta)) - c$, which equals zero for some $\eta'' < w_t$ due to the strict concavity of $u(\cdot)$. Hence, some $\bar{\eta} \in [\eta', \eta'']$ satisfies $F^P(f_t, w_t, \bar{\eta}) = 0$ for $\alpha \in (0, 1)$.

Finally, we need to show that $\bar{\eta}$ is unique. Because $F^R(\cdot)$ is an upper bound for $F^P(\cdot)$ for large η, it is sufficient to show that $F^P(\cdot)$ is quasi-convex in η. The first and second derivatives of $F^P(\cdot)$ with respect to η are

$$\frac{\partial F^P(f_t, w_t, \eta)}{\partial \eta} = \alpha h(\eta)\, (u(\eta) - u(w_t)) - (1 - H(\eta))\, u'(\eta)$$

and

$$\frac{\partial^2 F^P(f_t, w_t, \eta)}{\partial \eta^2} = \alpha h'(\eta)\left(u(\eta) - u(w_t)\right) + (1+\alpha)h(\eta)u'(\eta) - \left(1 - H(\eta)\right)u''(\eta).$$

Hence, if there is some $\hat{\eta}$ with $\partial F^P(f_t, w_t, \hat{\eta})/\partial \eta = 0$, we have

$$\frac{\partial^2 F^P(f_t, w_t, \hat{\eta})}{\partial \eta^2} = \frac{\left(1 - H(\hat{\eta})\right)h'(\hat{\eta}) + h(\hat{\eta})^2}{h(\hat{\eta})}u'(\hat{\eta}) + \alpha h(\hat{\eta})u'(\hat{\eta}) - \left(1 - H(\hat{\eta})\right)u''(\hat{\eta}) > 0,$$

since $\left(1 - H(\hat{\eta})\right)h'(\hat{\eta}) + h(\hat{\eta})^2 \geq 0$ by assumption 1.1 (ii). This completes the proof. $\square$

A.1.2 Proof of Hypothesis 1.1

Clearly, $\partial y/\partial w_t > 0$ if either $\partial n(w_t)/\partial w_t > 0$ or $\partial \bar{\eta}/\partial w_t > 0$. The second follows directly from application of the implicit function theorem on (1.2), iff $\alpha > 0$. $\square$

A.1.3 Proof of Hypothesis 1.2

Obvious, following the same argument as in the proof of Hypothesis 1.1. $\square$

A.2 Tables

Table A.1: Effect of Purchase-Date Weather on Ticket Orders for Repeat Customers

	Daily Ticket Orders		
	Fixed Effects Multiple Orders	Fixed Effects Multiple Orders per Season	Fixed Effects Multiple Orders with Bad Experience
Avg. Sun	0.012***	0.0080***	0.0037***
	(0.0028)	(0.0018)	(0.00089)
Avg. Rain	−0.0074***	−0.0031*	−0.00018
	(0.0028)	(0.0017)	(0.00080)
Forecast	Yes	Yes	Yes
Horizon Indicators	Yes	Yes	Yes
Observations	1635	1635	1635
Adjusted R^2	0.260	0.241	0.132

Notes: Coefficients and robust standard errors are reported for OLS regressions of daily ticket orders on purchase-date weather (sunshine duration in percent of time and rainfall in 1/100 mm), forecast of movie-date weather (forecasted temperatures and indicator variables for symbols), and horizon indicators (dummy variables for the number of days between purchase-date and movie-date). Fixed effects for the show are included. In the first column, the sample is restricted to sales to customers who have previously ordered tickets at least once between 2004 and 2011. In the second column, the sample is restricted to customers with multiple orders per season and in the final column to customers who had previously experienced rainfall during a show they had tickets for.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Predictive Power of Current Weather and Forecast

	Evening Sunshine Duration			
	1 Day Out	2 Days Out	3 Days Out	4 Days Out
Avg. Sun	0.032	0.023	−0.0027	0.077
	(0.051)	(0.050)	(0.050)	(0.052)
Avg. Rain	0.076	0.012	−0.040	−0.027
	(0.068)	(0.071)	(0.074)	(0.076)
Forecasted Maxtemp.	1.27**	0.91	1.39**	0.95
	(0.62)	(0.71)	(0.68)	(0.70)
Forecasted Mintemp.	−0.38	−0.026	0.097	0.13
	(0.79)	(0.93)	(0.93)	(0.85)
Symbol Partly Sunny	−20.6***	−27.0***	−16.3***	−19.8***
	(4.86)	(4.98)	(5.41)	(5.62)
Symbol Shower	−47.9***	−45.4***	−31.7***	−27.4***
	(5.10)	(5.36)	(5.42)	(5.58)
Symbol Rain	−43.2***	−43.0***	−36.3***	−25.2***
	(5.98)	(6.21)	(6.64)	(6.70)
Symbol T-Storm	−63.7***	−59.1***	−47.1***	−31.9**
	(9.04)	(10.1)	(12.0)	(14.0)
MA Rain 14 days	−0.11	0.27	0.29	0.11
	(0.27)	(0.28)	(0.29)	(0.29)
MA Sun 14 days	0.19**	0.23**	0.31***	0.25**
	(0.093)	(0.098)	(0.10)	(0.10)
Year and Month Indicators	*Yes*	*Yes*	*Yes*	*Yes*
Observations	470	469	470	471
Adjusted R^2	0.369	0.302	0.246	0.159

Notes: We report the coefficients and robust standard errors of OLS regressions estimating expected sunshine duration (in percent) in the evening between 5 pm and 7 pm based on information one to four days in advance (indicated in the column heading). The information comprises of current average sunshine duration, current rainfall (in 1/100 mm), the current weather forecast for the respective time horizon, a moving average of sunshine duration and rainfall of the past fortnight as well as year and month dummies.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Predictive Power of Current Weather without Forecast

	Evening Sunshine Duration			
	1 day out	2 days out	3 days out	4 days out
Avg. Sun	0.16***	0.043	0.024	0.052
	(0.049)	(0.049)	(0.050)	(0.050)
Avg. Rain	−0.064	−0.051	−0.071	−0.0047
	(0.065)	(0.066)	(0.066)	(0.066)
MA Sun 14 days	0.29***	0.35***	0.33***	0.27***
	(0.100)	(0.10)	(0.10)	(0.10)
MA Rain 14 days	−0.058	0.14	−0.00052	−0.12
	(0.28)	(0.29)	(0.29)	(0.29)
Year and Month Indicators	*Yes*	*Yes*	*Yes*	*Yes*
Observations	572	570	568	566
Adjusted R^2	0.130	0.099	0.088	0.082

Notes: We report the coefficients and robust standard errors of predictions of evening sunshine duration. The estimated models are identical to Table 8 with the exception that the variables of the weather forecast are excluded from the dependent variables.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Picture

Figure A.1: Location of the Theater

This picture shows the location of the theatre. Visitors are sitting in the amphitheatre on different rows on flaggings or on wooden boards (the area at the bottom left corner of the picture). The screen is on the left of this picture (not shown).

Appendix B

Appendix to Chapter 2

Table Appendix

The following tables provide estimates for five different sets of modern human resource controls, other firm related controls are unchanged.[1] Set 1 only includes a dummy variable indicating whether the respective firm uses competency tests. On top of that, Set 2 controls for personality tests of managers, whereas Set 3 additionally includes incentive payments. Set 4 and Set 5 are compound measures for the presence of modern human resource practises: The dummy in Set 4 equals one if either the firm uses competency tests or personality tests for managers or incentive pay. The indicator in Set 5 is one if all suggested measures, competency tests, personality tests for managers and incentive pay are present at the firm.

[1] We refer to the robustness section for an extensive discussion.

Table B.1: Robustness: Bottom Wage

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	-0.40^{**}	-0.38^{*}	-0.38^{*}	-0.38^{*}	-0.42^{**}
	(0.19)	(0.21)	(0.20)	(0.20)	(0.19)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2115	2115	2115	2115	2115
Adj. Wald Test					
F(2, 2153)		2.17	2.13		
Prob > F		0.12	0.12		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees earning bottom wages (below 4.5 pounds per hour) on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2: Robustness: Top Wage

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.16	0.023	-0.014	0.13	0.17
	(0.15)	(0.18)	(0.18)	(0.15)	(0.17)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2113	2113	2113	2113	2113
Adj. Wald Test					
F(2, 2153)		3.06	2.60		
Prob > F		0.05	0.07		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees earning high wages (above 15 pounds per hour) on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3: Robustness: Training

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.26*	0.25	0.25	0.25*	0.31**
	(0.14)	(0.15)	(0.15)	(0.13)	(0.14)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1888	1888	1888	1888	1888
Adj. Wald Test					
F(2, 2126)		1.95	1.90		
Prob > F		0.14	0.15		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of how many days employees are trained during one year on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Robustness: General Training

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.34*	0.23	0.23	0.32*	0.32*
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1964	1964	1964	1964	1964
Adj. Wald Test					
F(2, 2124)		3.36	3.22		
Prob > F		0.04	0.04		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of general training on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Robustness: Job Security

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.28	0.33*	0.34*	0.32*	0.34*
	(0.18)	(0.20)	(0.20)	(0.18)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	6982	6982	6982	6982	6982
Adj. Wald Test					
F(2, 2165)		1.43	1.51		
Prob > F		0.24	0.22		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of job security on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of all occupational group and includes a control for the number of occupational groups per firm. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Robustness: Benefits

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.37*	0.29	0.28	0.34*	0.43**
	(0.19)	(0.23)	(0.23)	(0.20)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		2.53	2.23		
Prob > F		0.08	0.11		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of benefits for the employees on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7: Robustness: No. of Benefits

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.30***	0.21	0.21	0.28**	0.28**
	(0.11)	(0.15)	(0.15)	(0.12)	(0.12)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		5.30	5.22		
Prob > F		0.01	0.01		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the number of provided benefits for the employees on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Robustness: Employer Pension Scheme

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.45***	0.38**	0.38**	0.47***	0.50***
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		3.47	3.44		
Prob > F		0.03	0.03		

Notes: We report the coefficients and robust standard errors of probit regressions of whether employer offer pension schemes on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Robustness: Extended Paid Leave

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.38**	0.28	0.27	0.35**	0.43***
	(0.16)	(0.18)	(0.19)	(0.16)	(0.17)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		3.54	3.09		
Prob > F		0.03	0.05		

Notes: We report the coefficients and robust standard errors of probit regressions of whether employer offer extended paid leave on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.10: Robustness: Team-Working

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.43***	0.37**	0.37**	0.46***	0.48***
	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2268	2268	2268	2268	2268
Adj. Wald Test					
F(2, 2187)		4.39	4.33		
Prob > F		0.01	0.01		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of what share of employees is designated to teams on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.11: Robustness: Monitoring

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.086	0.11	0.10	0.081	0.082
	(0.10)	(0.12)	(0.12)	(0.10)	(0.11)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2153)		0.39	0.36		
Prob > F		0.68	0.70		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the share of employees who have monitoring tasks on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.12: Robustness: Dismissals

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.0051	0.0024	0.0020	0.0039	0.0041
	(0.0059)	(0.0069)	(0.0069)	(0.0061)	(0.0062)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2152	2152	2152	2152	2152
Adj. Wald Test					
F(2, 2072)		0.67	0.54		
Prob > F		0.51	0.58		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees who have been dismissed during the previous year on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.13: Robustness: Firm Performance

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.28**	0.32*	0.31*	0.26*	0.23
	(0.14)	(0.17)	(0.17)	(0.14)	(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2148	2148	2148	2148	2148
Adj. Wald Test					
F(2, 2192)		1.98	1.79		
Prob > F		0.14	0.17		

Notes: We report the coefficients and robust standard errors of probit regressions of self-reported measure of firm performance being one if either managers report higher than median financial performance of their own firm, or higher labour productivity or higher product quality on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.14: Robustness: Firm Benefit

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.52***	0.63***	0.63***	0.52***	0.51***
	(0.14)	(0.16)	(0.16)	(0.14)	(0.14)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2189)		8.54	8.45		
Prob > F		0.0002	0.0002		

Notes: We report the coefficients and robust standard errors of probit regressions of compound measure of firm benefit being one, if the firm either uses higher than median team-working, less than median monitoring or reports higher than median firm performance as defined in Table B.13 on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.15: Robustness: Firm Benefit 2

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.27*	0.14	0.14	0.31**	0.23
	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2189)		2.38	2.59		
Prob > F		0.09	0.08		

Notes: We report the coefficients and robust standard errors of probit regressions of compound measure of firm benefit being one as defined in Table B.14 or has lower or equal to median turnover on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C

Appendix to Chapter 3

C.1 Tables

C.1.1 Data Description

Table C.1: Summary Statistics of Employee Characteristics

	Obs.	Avg.	SD	Min.	Max.
Age					
< 20	22362	0.03	0.18	0	1
20 - 29	22362	0.18	0.39	0	1
30 - 39	22362	0.25	0.43	0	1
40 - 49	22362	0.27	0.44	0	1
50 - 59	22362	0.22	0.41	0	1
60+	22362	0.05	0.21	0	1
Tenure					
< 1	22367	0.16	0.36	0	1
1 - 2	22367	0.13	0.33	0	1
2 - 5	22367	0.27	0.44	0	1
5 - 10	22367	0.19	0.39	0	1
10+	22367	0.26	0.44	0	1

Occupational Group					
Management	22762	0.11	0.31	0	1
Professional	22762	0.12	0.32	0	1
Associate Professional	22762	0.17	0.37	0	1
Administrative	22762	0.19	0.39	0	1
Skilled Trade	22762	0.07	0.25	0	1
Personal Service	22762	0.09	0.28	0	1
Sales	22762	0.07	0.26	0	1
Machine Operatives	22762	0.08	0.26	0	1
Routine and Unskilled	22762	0.11	0.32	0	1
Union					
Yes	22329	0.37	0.48	0	1
No, but in the past	22329	0.17	0.37	0	1
No, never	22329	0.47	0.50	0	1
Acad. Qual.	21991	2.03	1.21	1	8
Voc. Qual	21022	1.30	0.71	1	9
Gender	22345	0.54	0.5	0	1
Weekly Hours Working	22114	35.93	12.45	0	96

Notes: This table provides information on the number of observations, mean and standard deviation as well as minimum and maximum values of control variables for estimations on income and discretion. Statistics for each variable are calculated omitting answers "refusal", "don't know" and "not applicable", all indicating unclear answers.

Table C.2: Estimation of Income and Discretion Regressions

	Income	Discretion	
		What Tasks	How Work
	(1)	(2)	(3)
Occupational Group			
Professional	-8.12	0.33^{***}	0.21^{***}
	(8.87)	(0.028)	(0.023)
Associate	-85.3^{***}	0.30^{***}	0.19^{***}
	(8.25)	(0.027)	(0.022)
Secretary	-158.1^{***}	0.60^{***}	0.33^{***}
	(7.41)	(0.031)	(0.026)
Skilled Trade	-178.2^{***}	0.54^{***}	0.20^{***}
	(8.37)	(0.042)	(0.031)
Personal Service	-221.2^{***}	0.49^{***}	0.35^{***}
	(7.81)	(0.040)	(0.030)
Sales	-219.5^{***}	0.65^{***}	0.47^{***}
	(7.87)	(0.043)	(0.034)
Operatives	-233.0^{***}	0.79^{***}	0.52^{***}
	(8.74)	(0.042)	(0.037)
Unskilled	-264.2^{***}	0.62^{***}	0.37^{***}
	(8.26)	(0.036)	(0.030)
Age			
Age 18 - 19	-8.09	-0.088	-0.065
	(12.7)	(0.093)	(0.077)
Age 20 - 21	10.4	-0.14	-0.11
	(12.8)	(0.094)	(0.077)
Age 22 - 29	35.3^{***}	-0.15^{*}	-0.13^{*}
	(11.8)	(0.085)	(0.067)
Age 30 - 39	94.9^{***}	-0.28^{***}	-0.24^{***}
	(11.5)	(0.085)	(0.067)
Age 40 - 49	106.6^{***}	-0.30^{***}	-0.24^{***}
	(11.9)	(0.086)	(0.068)
Age 50 - 59	102.4^{***}	-0.30^{***}	-0.24^{***}
	(11.9)	(0.087)	(0.067)
Age 60 - 64	72.7^{***}	-0.38^{***}	-0.34^{***}
	(13.1)	(0.097)	(0.075)
Age 64+	16.2	-0.43^{***}	-0.26^{**}
	(21.2)	(0.13)	(0.11)

Tenure			
Tenure 1 - 2 years	4.88	-0.091^{***}	-0.00073
	(4.89)	(0.032)	(0.027)
Tenure 2 - 5 years	22.7^{***}	-0.12^{***}	-0.056^{**}
	(4.34)	(0.029)	(0.023)
Tenure 5 - 10 years	21.8^{***}	-0.17^{***}	-0.095^{***}
	(4.79)	(0.032)	(0.026)
Tenure > 10 years	43.3^{***}	-0.25^{***}	-0.14^{***}
	(5.23)	(0.031)	(0.025)
Gender	-78.0^{***}	0.0049	-0.041^{**}
	(3.82)	(0.019)	(0.016)
Constant	160.0^{***}	2.30^{***}	2.01^{***}
	(16.4)	(0.098)	(0.078)
Employee Controls	Yes	Yes	Yes
Observations	21506	21420	21428
R^2		0.087	0.059

Notes: This table provides estimation results for interval regressions of income (column (1)) and linear regression of discretion on employee observables. Predictions from these regressions are used to generate estimated values for income and discretion for each employee.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

C.1.2 Competency Tests

Table C.3: Regression of Lab. Prod. on HRM-Cluster – Competency Tests

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
HRM Cluster				
[0,0,0]		−0.31**		−0.33***
		(0.14)		(0.13)
[1,0,0]	−0.076	−0.39***	0.087	−0.25**
	(0.13)	(0.12)	(0.11)	(0.12)
[0,1,0]	0.097	−0.22	0.14	−0.19
	(0.16)	(0.15)	(0.15)	(0.15)
[0,0,1]	0.092	−0.22*	0.35***	0.015
	(0.14)	(0.13)	(0.11)	(0.12)
[1,1,0]	0.081	−0.23*	0.37***	0.035
	(0.15)	(0.14)	(0.14)	(0.15)
[1,0,1]	0.034	−0.28**	0.15	−0.19
	(0.14)	(0.14)	(0.13)	(0.14)
[0,1,1]	−0.041	−0.36**	0.15	−0.19
	(0.15)	(0.14)	(0.13)	(0.14)
[1,1,1]	0.31**		0.33***	
	(0.14)		(0.13)	
Union	−0.050	−0.050	−0.046	−0.046
	(0.085)	(0.085)	(0.084)	(0.084)
Pub. Sector	−0.065	−0.065	−0.070	−0.070
	(0.12)	(0.12)	(0.11)	(0.11)
Foreign	0.65***	0.65***	0.65***	0.65***
	(0.15)	(0.15)	(0.15)	(0.15)
Constant	3.69***	4.01***	3.52***	3.86***
	(0.20)	(0.21)	(0.19)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1812	1812	1812	1812
R^2	0.094	0.094	0.106	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster using competency tests as screening device and controls. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work". Columns (1) and (3) use cluster [0,0,0] as base category; column (2) and (4) omit cluster [1,1,1].

Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Regressions of Lab. Prod. on HRM Complementarities – Competency Tests

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
High Income (I)	0.11	0.097	0.19	0.14
	(0.17)	(0.16)	(0.16)	(0.15)
High Discretion (D)	0.13	0.092	0.39***	0.35***
	(0.14)	(0.14)	(0.12)	(0.11)
Comp. Test (C)	−0.083	−0.076	0.076	0.087
	(0.13)	(0.13)	(0.12)	(0.11)
I × D	−0.17	−0.23	−0.36*	−0.35*
	(0.21)	(0.20)	(0.20)	(0.19)
I × C	0.097	0.061	0.038	0.14
	(0.21)	(0.20)	(0.21)	(0.19)
D × C	0.045	0.018	−0.26	−0.29*
	(0.18)	(0.18)	(0.18)	(0.17)
I × D × C	0.18	0.35	0.32	0.25
	(0.28)	(0.27)	(0.28)	(0.27)
Union		−0.050		−0.046
		(0.085)		(0.084)
Pub. Sector		−0.065		−0.070
		(0.12)		(0.11)
Foreign		0.65***		0.65***
		(0.15)		(0.15)
Constant	3.46***	3.69***	3.33***	3.52***
	(0.11)	(0.20)	(0.095)	(0.19)
Firm Controls	*No*	*Yes*	*No*	*Yes*
Subpop. Observations	1815	1812	1815	1812
R^2	0.017	0.094	0.035	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, competency tests and its interactions. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work".

Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

C.1.3 Robustness

Table C.5: Labour Productivity on HRM-Cluster: Robustness I

	Raw		Categorical		Log Income	
	"What" (1)	"How" (2)	"What" (3)	"How" (4)	"What" (5)	"How" (6)
HRM Cluster						
[0,0,0]	−0.37**	−0.40***	−0.38***	−0.53***	−0.38***	−0.50***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
[1,0,0]	−0.27	−0.26	−0.31	−0.38**	−0.39*	−0.41**
	(0.22)	(0.20)	(0.20)	(0.19)	(0.21)	(0.21)
[0,1,0]	−0.36**	−0.29**	−0.22	−0.33**	−0.36**	−0.35**
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
[0,0,1]	−0.22	−0.23	−0.30**	−0.33**	−0.35**	−0.32**
	(0.15)	(0.15)	(0.14)	(0.15)	(0.14)	(0.15)
[1,1,0]	−0.66***	−0.64***	−0.43*	−0.61**	−0.43**	−0.53**
	(0.23)	(0.24)	(0.25)	(0.24)	(0.22)	(0.22)
[1,0,1]	−0.054	−0.10	−0.14	−0.24	−0.24	−0.29
	(0.17)	(0.19)	(0.18)	(0.19)	(0.17)	(0.18)
[0,1,1]	−0.23	−0.30**	−0.19	−0.28*	−0.27*	−0.32**
	(0.14)	(0.14)	(0.15)	(0.15)	(0.14)	(0.15)
Union	−0.048	−0.048	−0.053	−0.062	−0.058	−0.048
	(0.086)	(0.084)	(0.085)	(0.084)	(0.084)	(0.083)
Pub. Sector	−0.073	−0.063	−0.085	−0.070	−0.086	−0.058
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.58***	0.56***	0.56***	0.54***	0.57***	0.56***
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
Constant	3.97***	4.00***	4.01***	4.07***	4.07***	4.08***
	(0.22)	(0.23)	(0.23)	(0.23)	(0.22)	(0.23)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1498	1498	1495	1495	1494	1494
R^2	0.100	0.099	0.091	0.103	0.088	0.095

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls for different control specifications. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Labour Productivity on HRM Cluster: Robustness II

	Aggregation		Only Pers. Tests		Only Comp. Tests	
	"What" (1)	"How" (2)	"What" (3)	"How" (4)	"What" (5)	"How" (6)
HRM Cluster						
[0,0,0]	−0.41***	−0.43***	−0.50***	−0.50***	−0.31**	−0.33***
	(0.16)	(0.13)	(0.14)	(0.16)	(0.14)	(0.13)
[1,0,0]	−0.29	−0.31*	−0.40**	−0.36*	−0.39***	−0.25**
	(0.19)	(0.17)	(0.18)	(0.20)	(0.12)	(0.12)
[0,1,0]	−0.30*	−0.39***	−0.29**	−0.20	−0.22	−0.19
	(0.16)	(0.13)	(0.14)	(0.17)	(0.15)	(0.15)
[0,0,1]	−0.33**	−0.40***	−0.37***	−0.24	−0.22*	0.015
	(0.15)	(0.13)	(0.14)	(0.16)	(0.13)	(0.12)
[1,1,0]	−0.56**	−0.65***	−0.57**	−0.48*	−0.23*	0.035
	(0.24)	(0.22)	(0.26)	(0.28)	(0.14)	(0.15)
[1,0,1]	−0.23	−0.34*	−0.29*	−0.17	−0.28**	−0.19
	(0.20)	(0.18)	(0.17)	(0.19)	(0.14)	(0.14)
[0,1,1]	−0.31**	−0.33**	−0.37***	−0.33**	−0.36**	−0.19
	(0.15)	(0.14)	(0.14)	(0.16)	(0.14)	(0.14)
Union	−0.068	−0.061	−0.064	−0.042	−0.050	−0.046
	(0.084)	(0.084)	(0.084)	(0.084)	(0.085)	(0.084)
Pub. Sector	−0.078	−0.074	−0.068	−0.060	−0.065	−0.070
	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)
Foreign	0.57***	0.55***	0.55***	0.64***	0.65***	0.65***
	(0.13)	(0.13)	(0.12)	(0.13)	(0.15)	(0.15)
Constant	4.07***	4.12***	4.11***	4.02***	4.01***	3.86***
	(0.23)	(0.22)	(0.22)	(0.24)	(0.21)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1494	1494	1494	1494	1494	1494
R^2	0.089	0.091	0.093	0.103	0.094	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls for different control specifications. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: Labour Productivity on HRM Complementarities: Robustness I

	Raw		Categorical		Log Income	
	"What" (1)	"How" (2)	"What" (3)	"How" (4)	"What" (5)	"How" (6)
High Income (Inc.)	0.015	0.11	0.16	0.20*	0.025	0.15
	(0.11)	(0.10)	(0.099)	(0.10)	(0.11)	(0.10)
High Discretion (Disc.)	0.16	0.17*	0.081	0.20**	0.029	0.18*
	(0.11)	(0.10)	(0.099)	(0.094)	(0.11)	(0.095)
Pers. Test (PT)	0.11	0.15	0.073	0.14	−0.0065	0.091
	(0.19)	(0.16)	(0.17)	(0.15)	(0.18)	(0.17)
Inc. × Disc.	−0.027	−0.18	−0.053	−0.15	0.064	−0.15
	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
Inc. × PT	−0.41	−0.49*	−0.28	−0.43	−0.068	−0.27
	(0.28)	(0.26)	(0.28)	(0.26)	(0.27)	(0.26)
Disc. × PT	0.057	−0.021	0.090	−0.049	0.12	−0.064
	(0.23)	(0.22)	(0.22)	(0.21)	(0.22)	(0.22)
Inc. × Disc. × PT	0.47	0.67**	0.31	0.61*	0.22	0.57*
	(0.33)	(0.33)	(0.35)	(0.33)	(0.32)	(0.33)
Union	−0.048	−0.048	−0.053	−0.062	−0.058	−0.048
	(0.086)	(0.084)	(0.085)	(0.084)	(0.084)	(0.083)
Pub. Sector	−0.073	−0.063	−0.085	−0.070	−0.086	−0.058
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.58***	0.56***	0.56***	0.54***	0.57***	0.56***
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
Constant	3.60***	3.60***	3.62***	3.54***	3.68***	3.58***
	(0.21)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1498	1498	1495	1495	1494	1494
R^2	0.100	0.099	0.091	0.103	0.088	0.095

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests and its interactions. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".

Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table C.8: Labour Productivity on HRM Complementarities: Robustness II

	Aggregation		Only Pers. Tests		Only Comp. Tests	
	"What" (1)	"How" (2)	"What" (3)	"How" (4)	"What" (5)	"How" (6)
High Income (Inc.)	0.10	0.037	0.21**	0.30***	0.095	0.14
	(0.10)	(0.090)	(0.10)	(0.10)	(0.16)	(0.15)
High Discretion (Disc.)	0.082	0.035	0.13	0.25***	0.096	0.34***
	(0.10)	(0.10)	(0.11)	(0.092)	(0.14)	(0.11)
Pers. Test (PT)	0.12	0.12	0.032	0.11		
	(0.14)	(0.14)	(0.31)	(0.28)		
Comp. Test (CT)					−0.13	0.064
					(0.13)	(0.12)
Inc. × Disc.	−0.092	0.030	−0.21	−0.38***	−0.23	−0.34*
	(0.15)	(0.15)	(0.14)	(0.15)	(0.20)	(0.19)
Inc. × PT	−0.38	−0.38	−0.34	−0.42		
	(0.26)	(0.25)	(0.46)	(0.44)		
Inc. × CT					0.17	0.21
					(0.20)	(0.20)
Disc. × PT	−0.027	−0.065	0.18	0.073		
	(0.22)	(0.22)	(0.38)	(0.36)		
Disc. × CT					0.073	−0.29
					(0.20)	(0.20)
Inc. × Disc. × PT	0.59*	0.65**	0.75	0.84		
	(0.34)	(0.33)	(0.54)	(0.53)		
Inc. × Disc. × CT					0.27	0.22
					(0.29)	(0.29)
Union	−0.068	−0.061	−0.066	−0.047	−0.044	−0.041
	(0.084)	(0.084)	(0.089)	(0.090)	(0.091)	(0.091)
Pub. Sector	−0.078	−0.074	−0.088	−0.083	−0.087	−0.097
	(0.11)	(0.11)	(0.13)	(0.12)	(0.13)	(0.12)
Foreign	0.57***	0.55***	0.51***	0.61***	0.64***	0.66***
	(0.13)	(0.13)	(0.12)	(0.13)	(0.16)	(0.16)
Constant	3.66***	3.69***	3.62***	3.53***	3.70***	3.53***
	(0.20)	(0.21)	(0.22)	(0.21)	(0.21)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1494	1494	1108	1108	1108	1108
R^2	0.089	0.091	0.099	0.109	0.098	0.110

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests (competency tests for columns (5) and (6)) and its interactions. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

C.2 Figures

Figure C.1: Distribution of Employee Questionnaires per Firm

This figure provides relative frequencies of returned questionnaires per firm. Only firms with a minimum of one questionnaires are included.

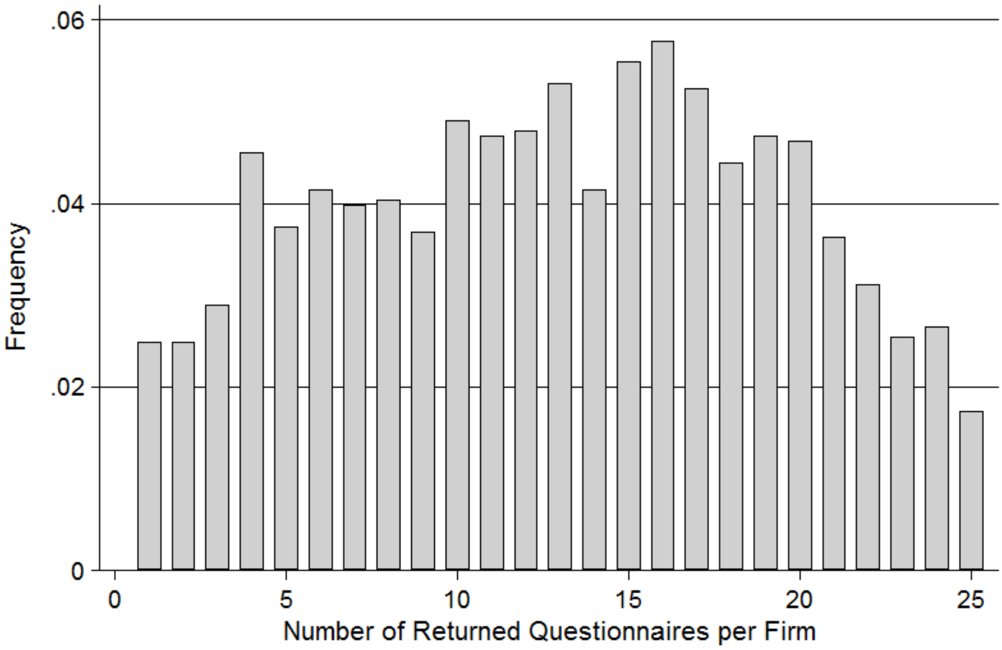


Figure C.2: Distribution of HRM Clusters

This figure provides relative frequencies of HRM clusters, which were used in Section 3.4. Panel (a) refers to "What Tasks" as measure of discretion, panel (b) to "How to Work".

(a) Discretion: “What Tasks” (b) Discretion: “How to Work”

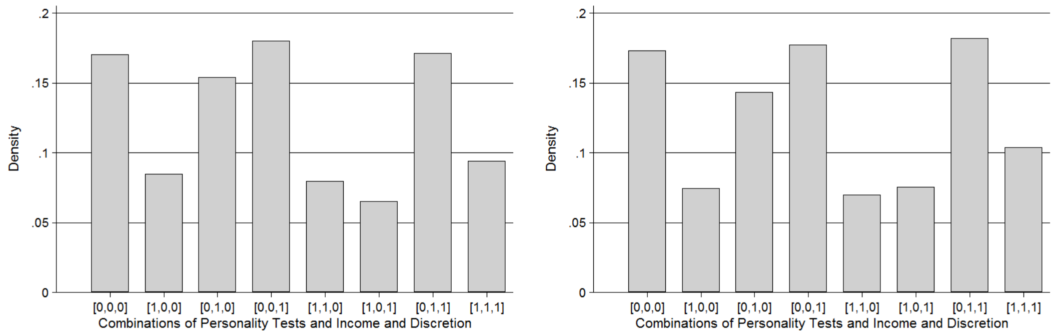


Figure C.2 depicts relative frequencies of each of the eight HRM clusters for two measures of discretion. The notation for each cluster is described as follows: [personality tests, high income, high discretion]. Each position is either 0 or 1 depending on whether firms require personality tests for job candidates, pays high income or allows for discretion. Hence on the very left of the figure cluster [0,0,0] describes firms which do not screen for personality, pay low income and do not grant high discretion and on the other extreme (cluster [1,1,1]) describes firms with personality tests, high income and discretion. Both panels exhibit similar frequency distributions. About 18 percent of firms belong to cluster [0,0,1], implying that these establishments do not screen job candidates for personality, do not pay higher than expected wages but provide substantial discretion. Approximately 10 percent of firms offer “good” jobs and screen for personality.

Bibliography

Akerlof, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97, 543–569.

Altmann, S., A. Falk, A. Grunewald, and D. Huffman (2013): "Contractual Incompleteness, Unemployment, and Labor Market Segregation," *mimeo*.

Arkes, H. R. and C. Blumer (1985): "The Psychology of Sunk Cost," *Organizational Behavior and Human Decision Processes*, 35, 124–140.

Ashton, M., S. Paunonen, E. Helmes, and D. N. Jackson (1998): "Kin Altruism, Reciprocal Altruism, and the Big Five Personality Factors," *Evolution and Human Behavior*, 19, 243–255.

Atkinson, R. L. and E. R. Hilgard (2000): *Hilgard's introduction to psychology*, Fort Worth and Tex: Harcourt College Publishers, 13 ed.

Autor, D. H. and D. Scarborough (2008): "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, 123, 219–277.

Baer, M. and M. Frese (2003): "Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance," *Journal of Organizational Behavior*, 24, 45–68.

Barber, B. M., T. Odean, and N. Zhu (2009): "Systematic noise," *Journal of Financial Markets*, 12, 547–569.

Bartling, B., E. Fehr, and K. M. Schmidt (2012a): "Screening, Competition, and Job Design: Economic Origins of Good Jobs," *American Economic Review*, 102, 834–864.

——— (2012b): "Use and Abuse of Authority: A Behavioral Foundation of the Employment Relation," *SSRN Electronic Journal*.

Becker, S., D. Messer, and S. Wolter (2011): "A Gift is Not Always a Gift: Gift Exchange in a Voucher Experiment," *CESifo Working Paper Series No. 3488.*

Bellemare, C. and B. Shearer (2009): "Gift giving and worker productivity: Evidence from a firm-level experiment," *Games and Economic Behavior*, 67, 233–244.

Ben-Ner, A. and A. Kramer (2011): "Personality and altruism in the dictator game: Relationship to giving to kin, collaborators, competitors, and neutrals," *Personality and Individual Differences*, 51, 216–221.

Ben-Ner, A., L. Putterman, F. Kong, and D. Magan (2004a): "Reciprocity in a two-part dictator game," *Journal of Economic Behavior & Organization*, 53, 333–352.

——— (2004b): "Reciprocity in a two-part dictator game," *Journal of Economic Behavior & Organization*, 53, 333–352.

Benartzi, S. (2001): "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," *The Journal of Finance*, 56, 1747–1764.

Black, S. E. and L. M. Lynch (2001): "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity," *Review of Economics and Statistics*, 83, 434–445.

Busse, M. R., D. G. Pope, J. C. Pope, and J. Silva-Risso (2012): "Projection Bias in the Car and Housing Markets," *NBER Working Paper*, 18212.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti (2010): "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts," *American Economic Review*, 100, 2261–2278.

Camerer, C., G. Loewenstein, and M. Rabin (2011): *Advances in Behavioral Economics*, Princeton University Press.

Charness, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22, 665–688.

Chavas, J.-P. (2000): "On Information and Market Dynamics: The Case of the U.S. Beef Market," *Journal of Economic Dynamics and Control*, 24, 833–853.

Choi, J., D. Laibson, B. Madrian, and A. Metrick (2009): "Reinforcement Learning and Savings Behavior," *The Journal of Finance*, 64, 2515–2534.

Cohn, A., E. Fehr, and L. Goette (2012): "Fair Wages and Effort: Evidence from a Field Experiement," *mimeo*, 2012.

Conlin, M., T. O'Donoghue, and T. J. Vogelsang (2007): "Projection Bias in Catalog Orders," *American Economic Review*, 97, 1217–1249.

Day, D. and S. Silverman (1989): "Personality and Job Performance: Evidence of Incremental Validity," *Personnel Psychology*, 42, 25–36.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2009): "Homo Reciprocans: Survey Evidence on Behavioural Outcomes," *The Economic Journal*, 119, 592–612.

Dur, R. and A. Glazer (2007): "Optimal Contracts When a Worker Envies His Boss," *Journal of Law, Economics, and Organization*, 24, 120–137.

Englmaier, F. and S. Leider (2012): "Contractual and Organizational Structure with Reciprocal Agents," *American Economic Journal: Microeconomics*, 4, 146–183.

Englmaier, F., S. Strasser, and J. Winter (2011): "Worker characteristics and wage differentials: Evidence from a gift-exchange experiment," *CESifo Working Paper Series No. 3637*.

Englmaier, F. and A. Wambach (2010): "Optimal incentive contracts under inequity aversion," *Games and Economic Behavior*, 69, 312–328.

Falk, A. and U. Fischbacher (2006): "A theory of reciprocity," *Games and Economic Behavior*, 54, 293–315.

Fehr, E. and S. Gächter (2000): "Fairness and Retaliation: the Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159–181.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108, 437–459.

Fehr, E. and K. Schmidt (2003): "Theories of Fairness and Reciprocity - Evidence and Economic Applications," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, 208 – 257.

Fehr, E. and K. M. Schmidt (2006): "Chapter 8 The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories," in *Handbook of the Economics of Giving, Altruism and Reciprocity, Volume I*, ed. by S.-C. Kolm and J. Mercier Ythier, Elsevier, vol. 1, 615–691.

Fuster, A., D. Laibson, and B. Mendel (2010): "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 24, 67–84.

Gneezy, U. and J. A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74, 1365–1384.

Güth, W., R. Schmittberger, and B. Schwarze (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3, 367–388.

Guthrie, J. (2001): "High-Involvement Work Practices, Turnover, and Productivity: Evidence from New Zealand," *Academy of Management Journal*, 44, 180–190.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2012): "Learning Through Noticing: Theory and Experimental Evidence in Farming," *NBER Working Paper*, 18401.

Hassan, T. A. and T. M. Mertens (2011): "The Social Cost of Near-Rational Investment," *NBER Working Paper*, 17027.

Hommes, C. (2011): "The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab," *Journal of Economic Dynamics and Control*, 35, 1–24.

Huang, F. and P. Cappelli (2010): "Applicant Screening and Performance-Related Outcomes," *American Economic Review*, 100, 214–218.

Ichniowski, C. and K. Shaw (2003): "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *The Journal of Economic Perspectives*, 17, 155–180.

Ichniowski, C., K. Shaw, and R. W. Crandall (1995): "Old Dogs and New Tricks: Determinants of the Adoption of Productivity- Enhancing Work Practices," *Brookings Papers on Economic Activity. Microeconomics*, 1995, 1.

Ichniowski, C., K. Shaw, and G. Prennushi (1999): "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87, 291–313.

Ichniowski, C. and K. P. G. Shaw (1997): "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87, 291–313.

Itoh, H. (2004): "Moral Hazard and Other-Regarding Preferences," *The Japanese Economic Review*, 55, 18–45.

Kaustia, M. and S. Knüpfer (2008): "Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions," *The Journal of Finance*, 63, 2679–2702.

Kube, S., M. A. Maréchal, and C. Puppe (2012): "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, 102, 1644–1662.

——— (2013): "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment," *Journal of the European Economic Association*, forthcoming.

Lemieux, T., W. B. MacLeod, and D. Parent (2009): "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 124, 1–49.

Leuven, E., H. van Oosterbeek, R. Sloof, and C. van Klavern (2005): "Worker Reciprocity and Employer Investment in Training," *Economica*, 72, 137–149.

Loewenstein, G. (1996): "Out of Control: Visceral Influences on Behavior," *Organizational Behavior and Human Decision Processes*, 65, 272–292.

Loewenstein, G., D. Nagin, and R. Paternoster (1997): "The Effect of Sexual Arousal on Expectations of Sexual Forcefulness," *Journal of Research in Crime and Delinquency*, 34, 443–473.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *The Quarterly Journal of Economics*, 118, 1209–1248.

Loewenstein, G. and D. Schkade (1999): "Wouldn't It Be Nice? Predicting Future Feelings," in *Well-Being: The Foundations of Hedonic Psychology*, ed. by D. Kahneman, E. Diener, and N. Schwarz, New York: Russell Sage Foundation.

Machin, S. and M. Stewart (1996): "Trade Unions and Financial Performance," *Oxford Economic Papers*, 48, 213–241.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and J. T. Lowell (2002): "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment," *American Economic Review*, 92, 850–873.

Nordgren, L. F., J. van der Pligt, and F. van Harreveld (2007): "Evaluating Eve: Visceral States Influence the Evaluation of Impulsive Behavior," *Journal of Personality and Social Psychology*, 93, 75–84.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

Read, D. and B. van Leeuwen (1998): "Predicting Hunger: The Effects of Appetite and Delay on Choice," *Organizational Behavior and Human Decision Processes*, 76, 189–205.

Rynes, S. and D. Cable (2003): "Recruitment Research in the Twenty-First Century," in *Handbook of Psychology*, 55–76.

Schwartzstein, J. (2012): "Selective Attention and Learning," *Working Paper*.

Simonsohn, U. (2010): "Weather to Go to College," *Economic Journal*, 120, 270–280.

StataCorp. (2011): "Stata 12 Base Reference Manual," *College Station, TX: Stata Press.*

Thaler, R. (1988): "Anomalities The Ultimatum Game," *Journal of Economic Perspectives*, 2, 195–206.

van Boven, L. and G. Loewenstein (2003): "Social Projection of Transient Drive States," *Personality and Social Psychology Bulletin*, 29, 1159–1168.

Wall, T. D., J. Michie, M. Patterson, S. J. Wood, M. Sheehan, C. W. Clegg, and West Michael (2004): "On the Validity of Subjective Measures on Company Performance," *Personnel Psychology*, 57, 95–118.

Wilk, S. and P. Cappelli (2003): "Understanding the Determinants of Employer Use of Selection Methods," *Personnel Psychology*, 56, 103–124.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

20. Juni 2013

Thomas Kolaska