INCENTIVE PAY AND PERFORMANCE:

INSIDER ECONOMETRICS IN A MULTI-UNIT FIRM[*†]

Hein Bogaard
George Washington University

Jan Svejnar
Columbia University, CERGE-EI, CEPR, IZA

December 2013

Abstract

We exploit organizational reforms in a foreign-owned bank in Central-East Europe to study the implementation of modern HRM policies in an emerging market context. We have branch-level data and use our knowledge of the process that led to the adoption of the reforms to implement two estimators that address endogeneity bias in a complementary fashion: an IV approach and Generalized Propensity Score estimation. Our results show that some of the reforms had a positive impact on productivity, but they also underscore the risks of quantity-based incentives where quality is important.

Keywords: Insider Econometrics, Endogeneity of HRM Policies, Incentives, Foreign Ownership, Banking, Central and Eastern Europe,

JEL Classification: F23, G21, M52

[*] Hein Bogaard, School of Business, George Washington University, 401 Funger Hall, 2201 G Street NW, 20052 Washington D.C. phone: + 1 202 994 0820, fax: +1 202 994 7422, email: hbogaard@gwu.edu
Jan Svejnar, School of International and Public Affairs, Columbia University, 1508 International Affairs, MC3328, 420 West 118th Street, New York, NY 10027, phone: +1 212 854 0135, fax : +1 212 854 1226, email: svejnar@columbia.edu
[†] We thank "The Bank" for data access and co-operation in analyzing the data and Štěpán Jurajda, Yuriy Gorodnichenko, Mario Macis and Jordan Siegel as well as participants at the Academy of International Business, the CESIfo conference on Banking and Institutions, the 2010 IZA World Bank conference in Cape town, the 16th Dubrovnik Economic conference, the 2011 meetings of the Association for Comparative Economic Studies and seminars at the EBRD and Tilburg University for comments on previous drafts. Bogaard acknowledges support from the Jean Monnet fellowship of the European Union Institute at the University of Michigan. Svejnar benefitted from a research grant from the World Bank and Grant Agency of the Czech Republic (Grant P402/10/2130).

I. Introduction

As firms in emerging markets are confronted with rising wages and institutional reforms there has been a growing interest in the efficacy of modern HRM policies, as well as of modern management and of information technology, in these markets (Commander, Harrison and Menezes-Filho 2011; Bloom, Eifert, Mahajan, McKenzie and Roberts 2013). In this paper, we analyze an episode of organizational reform at a foreign-owned bank in Central-East Europe to expand our understanding of the relationship between HRM policies and performance in an emerging market context. Our analysis uses data from 180 branches over 20 quarters between 2003 and 2007. Over this period, the bank rolled out new HRM policies in its branches, which introduced job differentiation and high-powered incentives for sales staff (but not for cashiers).

The extent to which incentives improve performance has been an important topic in the personnel economics literature (Lazear 2000). Most of the literature studies the use of incentives for workers or teams with relatively homogeneous tasks. In our setting, workers have heterogeneous, but complementary, tasks. Designing optimal incentives in such a setting is complicated. The system introduced by the bank is consistent with the notion that workers with tasks that contribute to the bottom line should receive high-powered incentives, while multi-taskers should not (Holmstrom and Milgrom 1991; Besanko, Regibeau and Rockett 2005). However, differentiation of incentives has the potential to induce collusion between workers and more generally, quantity-based incentives carry risk in a setting such as banking where quality is important (Baker 2002).

Our empirical approach that is grounded in the insider econometrics literature, a branch of personnel economics that has sought to use detailed knowledge of the firm to analyze the impact of modern HRM policies (Ichniowski and Shaw 2003; 2012). Policies of interest have included incentive schemes, team work and task flexibility. To the extent that insider econometric studies rely on non-experimental data endogeneity bias has been an important methodological concern. Bias arises due to unobserved differences in (i) the performance of organizational units or (ii) the marginal benefit of HRM practices (which can be due to complementarities between practices). The first source of bias can be addressed with fixed effects estimation and some authors show that fixed effects effectively deal with all plausible sources of bias (Ichniowski, Shaw and Prennushi 1997; Athey and Stern 2002). However, fixed effects estimation is not generally valid.

In this paper, we use our knowledge of the process that led to the adoption of HRM reforms to develop and instrumental variables approach to estimating the impact of the HRM reforms on branch-level performance. Our IV approach makes exploits the fact that in our data, all branches were subject to a common set of policy decisions at headquarters even though HRM reforms were implemented at different times and to a different extent across branches. Hence, for each branch, we can use information on the implementation of HRM reforms in branches that are observationally similar to construct instruments that approximate the strength of the policy shock, but are independent of the branch-specific benefits of the policies. While our dataset is somewhat unique in the sense that it comprises the universe of organizational units that are eligible for the HRM reforms that we study our approach should be valid in other datasets. Specifically, our method can be applied if organizational units (be they independent firms or affiliated to a single one) are subject to policy shocks that are exogenous at the level of observation and affect the propensity that an HRM policy will be introduced.

Using our IV approach we find that the introduction of sales staff with high-powered individual incentives contributes to the average sales productivity of branch employees. This effect is larger in large branches, where free-riding is a problem if there are no individual incentives, but declines when the ratio of sales staff to other staff, who provide administrative and other services that are complementary to sales effort, becomes too large. However, we do not find that the HRM policies improved profitability, the product mix or loan quality. Overall, the results point to the effectiveness of the new organizational structure and bonus system in eliciting effort from branch staff. At the same time, the results raise some concerns about the effect of differentiation in incentives.

Our results are very robust within the context of IV estimation, which controls for *unobserved* heterogeneity by replacing, in the second stage, the actual adoption of HRM reforms by the propensity of a branch to adopt the reforms. A weakness of this approach is that results may reflect structural differences between branches that correlate with the propensity for treatment rather than actual treatment (Blundell and Costa Dias 2009). We therefore compare our IV estimates to Generalized Propensity Score (GPS) estimates (Imbens 2000; Flores and Mitnik 2013). These estimates control for selection on *observable* differences across branches. They represent estimates of the impact of the HRM reforms on branch productivity that are based on comparisons between branches with the same propensity for treatment but different actual treatment. The strength of the GPS estimates lies precisely in the area of weakness of the IV estimates and if the results from both estimators are similar (as they do in our case), they reinforce each other.

In what follows we first discuss the bank and our data (section II). We next discuss related literature an empirical predictions (section III). Subsequently, we present our IV approach (section IV) and findings (section V). We present further robustness tests in the form of GPS estimation in section VI and we conclude in section 0.

II. Background and Data

Banking in the CEE region has changed dramatically since the early 1990s when there were primarily universal, state-owned banks that suffered from an overhang of bad debts and were known for poor management and service (Buch 1997; Berglof and Bolton 2002). Today, all countries in the region have a modern banking sector with relatively well-managed banks with foreign ownership and a range of client-friendly products on offer.

The bank that we study is a leading financial institution in its market and has over 200 branches that serve retail and SME clients. Upon privatization in the late 1990s, a majority of shares were acquired by a West-European bank, which later purchased the remaining shares.

We have access to quarterly branch-level balance sheets and profit and loss statements covering the five-year period from 2003 to 2007. The data include a quarterly overview of staff for each branch, broken down by functions. The objective of the branches is to maximize the "sales" of deposits, loans and savings products to retail and SME clients. In the context of this paper, it is appropriate to think of branches as "outlets" rather than "mini-banks". A branch's ability to lend is restricted by rules related to the assessment of creditworthiness of borrowers but not by its intake of deposits – the loan-deposit balance is monitored at the bank level.

Following acquisition, the foreign owner introduced a range of organizational reforms, initially to improve governance, risk management and cost-effectiveness. We focus on the second phase of reforms during which the bank sought to improve the commercial orientation and client focus of the branches. The reforms had three key elements. First, there was an effort to improve client segmentation into high-value and regular clients. Second, mirroring the segmentation of clients, the bank created a new functional structure. Under the old structure, there were differences in seniority, but function profiles were otherwise ill-defined (figure 1). In

2003 the bank introduced “bankers,” who focus on high-value clients within either the retail or SME market. Two years later, the bank created “advisor” functions. Like bankers, advisors are expected to focus on sales and client relations, but instead of engaging with specific clients, they specialize in specific products such as mortgages and contractual savings. The banker and advisor functions were not created all at once and we use variation in the number of bankers and advisors over time and across branches to evaluate the impact of the reforms.

With the introduction of the banker and advisor functions, remaining branch staff was expected to focus on administrative and transactional services as well as sales of regular products. The bank created specific function profiles and training programs for both bankers and advisors and although most of the recruitment for these functions happened within branches, the process was perceived as a clear shift in the valuation of skills.

The third key element of the branch-level reforms involved the introduction of a new incentive system. Before 2003, performance bonuses put a significant weight on branch profits, which are far removed from branch employees' day-to-day activities. Under the new system, bonuses are largely based on performance towards sales targets. Regular branch staff receive a bonus of 10 percent of their regular salary if the branch as a whole meets 70 percent of its target. The maximum bonus is 40 percent of salary if the branch hits 200 percent of target performance. Advisors have the same bonus curve, but their performance is measured on the basis of a 70/30 weighted average of progress towards individual sales targets and branch-level sales targets. The performance of bankers is also measured as a 70/30 weighted average of individual and branch performance. However, the bonus curve of bankers is steeper. Bankers can receive a maximum bonus of 75 percent of their regular salary when they reach 150 percent of target sales.

To determine sales targets, the bank uses an econometric model to estimate the sales potential of each branch on the basis of local economic variables and sales experience in the region. The sales performance of any branch has only a small impact on the central tendency in the regression line that establishes future sales targets. This limits ratchet effects and strategic behavior to influence targets (Weitzman 1980; Murphy 2000). Low performance in the current period leads to an immediate drop in bonuses, but not to lower future lending targets.[1]

In our sample period, the bank also conducted a variety of training programs for both new and existing employees, including an executive education program for branch managers. We control for this "Leadership Academy" in our empirical analysis, but this program was introduced too late into our sample period to expect a measurable impact on performance. The impact of other training programs, as well as the introduction of the service standards, is absorbed in time fixed effects.

Table 1 gives an overview of our data and in particular of changes in the functional structure in the branches. In panel A, the branches are divided in three groups by size. The smallest branches focus exclusively on retail clients and even by the end of the sample period, they have only a limited number of bankers and advisors. Indeed, in 2007, the bank decided to discontinue some of the advisor positions in these branches because it felt that they were not sufficiently productive relative to cost. In medium and especially in large branches, the number of bankers per employee increases in 2003 and 2004 and then stabilizes. The same happens with advisors in 2005 and 2006. Large branches have a stronger focus on SME clients and also a larger share of SME bankers per employee. Panel B of the table shows that the appointment of retail and SME bankers is associated with higher loan growth and higher profitability.

---

[1] The regression approach did not work to the bank's satisfaction for SME products. Targets for SME loans and deposits are based on assumptions about achievable sales per employee.

III. Related Literature and Empirical Predictions

The new organizational structure of the branches, while probably not uncommon, is complex in the sense that it involves heterogeneity of functions, complementarity between tasks and differentiated incentives. As a result, it is difficult to judge whether the incentive system is optimal given the roles of bankers, advisors and other employees and vice versa (Besanko, Regibeau and Rockett 2005; Corts 2007). Nevertheless, the literature provides us with substantial insight into the likely impact of the reforms on branch performance.

Theory suggests that the new bonus system will improve sales productivity overall and in particular in branches with banker or advisor positions. In terms of the principal-agent model, the bonus system strengthens the link between effort directed at sales and the signal (sales rather than profits) that is used to determine the bonus. For bankers and advisors, individual sales targets should also reduce free riding – this is especially important in large branches where the incentive to free ride is largest in the absence of individual targets (Alchian and Demsetz 1972).

In combination with the bonus system, the redefinition of function profiles ought to enhance branch performance further. First, Lazear (2000) shows that high-powered incentives not only induce more effort, but also improve performance because they attract workers who are more productive. Within the bank, the creation of banker and advisor positions facilitated matching of employees to functions. While we do not have data on individuals, we were told that several cashiers became very successful in banker positions while some former branch managers moved into support roles and not be subject to high-powered incentives. Second, the incentive structure is aligned with the view that bankers and advisors should focus on making sales, whereas administrative staff and cashiers are multi-taskers who make sales but also provide support services (Holmstrom and Milgrom 1991). Specifically, Besanko, Regibeau and Rockett

(2005) argue that a "functional" organization with function specific reward schedules becomes more desirable if one function (e.g., sales) makes a higher marginal contribution to performance than another (e.g., support services) and if certain activities generate externalities (cashiers serve both retail and SME customers and support performance in both product segments).

Beyond the main effects of the reforms on employee effort and matching, the literature also suggests potential drawbacks. First, administrative and transaction services performed by cashiers are complementary to the sales effort by bankers and advisors. Corts (2007) argues that this arrangement may lead to under-provision of effort by workers who are not rewarded for an important output. This is especially true if workers lack "intrinsic motivation" which they are assumed to have by Holmstrom and Milgrom (1991). In our empirical context, effective delivery of service effort by cashiers is especially important if there is a high number of bankers and advisors relative to the number of cashiers. We therefore expect that the impact of additional bankers on branch productivity eventually declines with the number of bankers per cashier.

Another drawback of the differentiation in incentives is that it invites collusion among branch employees (Tirole 1986; Laffont and Rochet 1997). To branch employees collectively, a sale is worth more if it is made by a banker than by a cashier because a banker receives a higher bonus at the margin. Hence, if a cashier is about to make a sale it is in the interest of both bankers and cashiers to exchange a bribe and record the sale as being made by the banker. Such bribery is not merely a theoretical possibility; in the past, the bank allowed agents to sell some of its products on a commission basis. The bank suspended this practice when it was found that agents bribed branch employees into letting them book sales that would have been made anyway.

The existence of bribery between bankers and other employees would not necessarily affect the volume of sales, but it would reduce profitability. More generally, the literature has found that quantity-based incentives tend to have a limited impact on profits because they are too expensive (Cappelli and Neumark 2001; Freeman and Kleiner 2005), or because they encourage lending to bad risks (Baker 2002; Agarwal and Wang 2009).

By way of summary, we have four empirical predictions. First, the introduction of bankers and advisors and the associated bonus system should lead to an increase in sales productivity as it encourages effort and improves matching of employees to jobs. Second, the positive effect of bankers and advisors should be higher in large branches because without individualized incentives free-riding is more prevalent in those branches. Third, the positive impact of bankers and advisors should decline when their number gets too high relative to the number of cashiers because cashiers may not provide sufficient service effort to support bankers and advisors. Fourth, the impact of the introduction of bankers and advisors on branch-level profitability is smaller than the impact on sales because differentiation in incentives encourages bribery and because quantity based incentives reduce attention to quality.

IV. Empirical Strategy

Our empirical approach builds on and extends the insider econometrics literature (Ichniowski and Shaw 2003; 2012). This literature has sought to use detailed knowledge of organizations to evaluate the impact of HRM policies on performance. While some research on HRM and other managerial practices has used experiments (Wageman 1995; Bloom, Eifert, Mahajan, McKenzie and Roberts 2013), many researchers have exploited access to data that was collected by firms in the course of doing business. Such data are often rich and detailed and they have contributed to important insights into the effectiveness of e.g. incentives on performance (Lazear 2000;

Freeman and Kleiner 2005). However, the non-experimental nature of the data also has the potential to cause endogeneity bias in estimates of the impact of HRM policies on performance.

To evaluate HRM policies, the researchers have generally relied on models that relate output or productivity to firm characteristics and then augmented these models with indicators of HRM policies:

$$Y_{it} = \alpha + X_{it}\beta + W_{it}\gamma + \varepsilon_{it} \qquad (1)$$

In this equation, Y is output or productivity, X is a vector of HRM policies and W is a vector of unit characteristics and control variables and ε is a mean zero error term. Estimates of β are subject to endogeneity bias if HRM reforms are more likely to be implemented in organizational units i that (i) perform systematically better or worse than other units or (ii) where the marginal impact of reforms is be higher (Athey and Stern 1998). The first source of endogeneity can be differenced out, but the second source cannot, which is most easily illustrated with a decomposition of the error term ε in equation (1):

$$Y_{it} = \alpha + X_{it}\beta + W_{it}\gamma + \mu_i + X_{it}\nu_i + \omega_{it} \qquad (2)$$

In equation (2), unit specific performance is represented by μ_i, which disappears in a fixed effects specification so that any bias that arises due to correlation between X and μ disappears. By contrast, fixed effects do not eliminate ν_i, the branch-specific contribution of X to productivity. For each unit, the marginal contribution of X to productivity is the average productivity of X, the coefficient β, plus the branch specific contribution ν_i. First differencing leaves $\nu_i(X_{it} - X_{it-1})$ in the error term. If the adoption of X_{it} is optimal, the reform is more likely to be adopted where ν_i is high, so that $(X_{it} - X_{it-1})$ and ν_i are positively correlated. The result is upward bias in estimates of β in both OLS and mean- or first-difference estimates (note that this is exactly what Lazear

(2000) shows; he finds that the positive impact of incentives is partially due to self-selection of more productive workers into a regime with high-powered incentives).

In randomized trials the existence of v_i is not a problem because randomization eliminates correlation between X_{it} and v_i. In non-experimental data, the solution to heterogeneity in the benefits of adoption is context specific. For example, Ichniowski, Shaw and Prennushi (1997) argue that adoption of the HRM policies they test is a function of differences in the cost of adoption, not the benefits (i.e. they argue that v_i is equal to zero). Athey and Stern (2002) and Bartel, Freeman, Ichniowski and Kleiner (2011) show that, in their data, fixed effects eliminate all but very implausible sources of bias. There is no basis for such arguments in our empirical context and we therefore develop an IV approach that exploits our knowledge of the process that led to the adoption of the new HRM policies in the branches of the bank.

In our data, some branches have more bankers and advisors than others, even if they are of the same size. However, all appointments are the result of a policy shock that results from a strategic decision at headquarters and is exogenous to all branches. In particular, assume that branch i belongs to a group of K branches that are observationally similar because they belong to the same size-class or are located in the same region. At any point in time, the number of bankers and advisors in all branches $k \neq i$, where $k, i \in K$, is representative of the policy impulse coming from the bank's headquarters that branch i receives, but it is uncorrelated with v_{ij}, the branch-specific contribution to productivity of bankers and advisors in branch i. The creation of banker and advisor positions in branches $k \neq i$ is based on sales prospects at these branches, but it does not depend on prospects for branch i. Therefore, we can use information on the implementation of the reforms in branches $k \neq i$ as instruments to identify the exogenous component of the reforms in (Hausman and Taylor 1981; Hausman 1997; Shirley and Xu 2001).

A. Empirical Model

The principal production factor of the branches in the bank is labor and the function underlying our empirical model posits that output is a function of the number of employees in a branch. Branch output is measured as the sum of deposits and loans a branch makes, which the branches are incentivized to maximize. Following Bartel, Freeman, Ichniowski and Kleiner (2011), we call the sum of deposits and loans "footing".[2] To facilitate the interpretation of results in terms of productivity, we use net sales per employee, $\Delta Footing_{ijt} / FTE_{ijt}$, as the dependent variable, with *FTE* for Full-Time Equivalent and the indices by *i*, *j* and *t*, stand for branch, region and time.

We measure the implementation of reforms as the number of bankers plus advisors in a branch divided by the number of employees. Because we anticipate that the impact of bankers and advisors on productivity will be higher in large branches and that it will decline as the share of bankers and advisors increases, we include the main effect of the reforms in our model along with a squared effect and an interaction with branch size.

$$\begin{aligned}\frac{\Delta Footing_{ijt}}{FTE_{ijt}} &= \gamma + \theta_1 \frac{(Bankers + Advisors)_{ijt}}{FTE_{ijt}} + \theta_2 \left(\frac{(Bankers + Advisors)_{ijt}}{FTE_{ijt}} \right)^2 \\ &+ \theta_3 \frac{(Bankers + Advisors)_{ijt}}{FTE_{ijt}} \times FTE_{ijt} + \beta_1 FTE_{ijt} + \beta_2 FTE_{ijt}^2 + controls + \varepsilon_{ijt}\end{aligned} \tag{3}$$

The coefficients of interest are the main effect of the reforms θ_1, which we expect to be positive, the squared effect of the reforms θ_2, which we expect to be negative and the interaction with the number of employees θ_3, which we expect to be positive.[3]

---

[2] In the banking literature, footing is aligned with the "production approach" which holds that both lending and deposit taking are services that banks provide to their clients and should be counted as outputs (Berger, Hanweck and Humphrey 1987). The alternative is the intermediation approach, which claims that banks produce assets use deposits as inputs (Sealey and Lindley 1977). The intermediation approach has merit at the bank level, but not at the level of the branches since branch lending is not constrained by the availability or cost of deposits.

[3] Note that the interaction $(Bankers + Advisors/FTE) \times FTE$ reduces to $Bankers + Advisors$, which we use going forward.

We treat the number of bankers and advisors per employee, its square and its interaction with the number of employees as endogenous and construct instruments for these variables following the approach we sketched above. Specifically, we use the following instruments: the average number of employees in branches in the same region, the average number of retail bankers in branches in the same size class, the share of branches in the same region with at least one SME banker, the share of branches in the same size class with at least one advisor, and a categorical variable (ranging from 1 to 4) indicating progress with the rollout of the program that introduced the bankers. Hence, for each quarter and for each branch i, our instruments are the average branch characteristics calculated across all branches $k \neq i$ in K where K is defined by region or by size class (Table 1).

In addition to the variables listed in equation (3) we control for demand conditions with the municipal unemployment rate and two dummies indicating the size of the municipality in which a bank is located (population between 50,000 and 100,000, or population > 100,000; the capital, which is the largest city in the country, is a separate region in the bank's organization). In addition, we include a dummy that is equal to 1 when a branch manager has participated in the Leadership Academy and 0 otherwise. Finally, we control for time and location with a full set of region × time fixed effects.

We estimate our models in Stata using GMM, implemented with the *ivreg2* command (Baum, Schaffer and Stillman 2007). We report Hansen's J-test to show that the instruments can be omitted from the main equation,[4] and first-stage F-tests and the Kleibergen-Paap test to check that the first-stage regressions do not suffer from underidentification.

[4] The null hypothesis of the J-test is that the excluded instruments have no explanatory power in the main equation. Therefore, if we reject the null hypothesis, the instruments are not valid.

V. Results

In Table 2 we report the estimated coefficients from the baseline model using both OLS and GMM estimation. The difference between the two GMM estimates is that the model in column 2 treats only Bankers + Advisors / FTE, its square and Bankers + Advisors as endogenous whereas the model in column 3 also treats FTE and its square as endogenous. The estimations generate several interesting results. First, the main effect of (*Bankers + Advisors*) / *FTE* is positive. The coefficient is larger in the GMM estimates (column 2 and 3) than in the OLS estimates (column 1). Endogeneity causes a downward bias in the OLS estimates, suggesting that bankers and advisors were assigned to branches that initially had relatively low productivity. Second, in the GMM estimates, the coefficient on the squared term is significantly negative. This implies that the relationship between the ratio of bankers and advisors to FTE and sales productivity is concave as predicted (the inflection point lies around 0.15). Third, recall that *Bankers + Advisors* is equal to ((*Bankers + Advisors*) / *FTE*) × *FTE* (footnote 3). The positive coefficient on Bankers + Advisors therefore implies that the impact of the HRM reforms is higher for large branches than for small ones. At the same time, the negative coefficient on FTE implies that sales productivity is lower in large branches on average. Together, these results are consistent with the prediction that free riding under group incentives is more problematic in larger groups (in large branches individual incentives "solve" a bigger problem).

Considering that the marginal contribution of bankers and advisors to sales productivity depends on branch size, we calculate point estimates of this contribution for each of the branches. The GMM estimates in column 2 yield positive and significant marginal contribution in about 55 percent of the observations. It is negative and significant in fewer than 20 percent of the observations from branches that have a relatively high number of bankers and advisors per employee. On average, the marginal contribution is about one-and-a-half standard deviations of the quarterly increase in footing per employee. Hence, although some branches appear to have too many bankers and advisors, their overall contribution to sales productivity is positive.

The first-stage F-tests suggest that the instruments are sufficiently strong and the J-test implies that the omitted instruments have no explanatory power in the main regression. Furthermore, a "Difference-in-J" test suggests that the GMM regression in column 2 is significantly different from the OLS regression and that the ratio of bankers and advisors to employees should be treated as endogenous. In contrast, using the same test, there is no evidence that the regression in column 3, which treated FTE and its square as endogenous, is different from the one in column 2. Although we should not interpret this as evidence that FTE is exogenous, we use the specification in column 2 as our baseline regression.

In unreported regressions, we also estimate a model without the squared and interaction terms and we find that the full specification in Table 2 fits the data better. Also, we estimate a model that includes operational expenses at the branch level in addition to the number of employees as a control variable. This model produces almost identical results.

A. Additional evidence

Building on the result that giving a subset of branch employees high-powered incentives raises sales, we perform additional analysis to ascertain the robustness of our findings. We estimate the model while excluding the regions one-by-one to ensure that none of the regions or branches dominates the results.[5] None does. Similarly, we estimate the model with the years eliminated one-by-one. Again, the results are robust. We also estimate a model in which we include the

[5] In some of the regressions, the coefficient on *Bankers + Advisors* is not significant at conventional levels. However, the p-value is generally close to 10%, just like the p-value in Table 2.

members of the banker teams (assistants and managers) in the count of employees with high-powered incentives. Over the course of our sample period, banker teams were formalized in the branch organization and the incentives for the members became more closely aligned with those of the bankers. Again, the results remain unchanged. Finally, if there is positive correlation between *Bankers + Advisors* and branch-specific productivity of bankers and advisors, v_i in equation (2), there is, in theory, some negative correlation between the instrumental variables and branch-specific productivity. The validity of our instrumental variables is based on the assumption that the sample is large enough to ignore this correlation. Hansen's J-test suggests this is so. As another check, we estimate our model with the Jackknife Instrumental Variables Estimator (JIVE, Angrist, Imbens and Krueger 1999). To eliminate correlation between v_{ij} and the instrumented variables in the first stage, JIVE excludes both the instrumental and instrumented variable for observation *i* from the estimation of the first-stage equation for observation *i*. The JIVE estimates are almost identical to those in Table 2.

B. Profitability and Quality

We anticipate that the introduction of bankers and advisors and quantity-based incentives will have a smaller impact on profitability than on sales volume and that it may also affect indicators of quality such as the volume of bad loans and loan-loss provisions. In addition, we investigate whether the reforms affect the product mix. Bankers and advisors were expected to raise sales of mortgages and contractual savings products with the aim to tie clients to the bank long term.

Regression results are reported in Table 3. In columns 1 and 3 of the table, we find that bankers and advisors contributed to higher sales per employee of mortgages and mutual fund type products (due to data availability, the regressions in columns 1 to 4 only cover the years

2005 to 2007). However, this did not translate into a larger share of these products in loans and deposits outstanding (columns 2 and 4). In columns 5 and 6 of Table 3, we investigate the impact of the reforms on loan quality. In large branches, there is a positive relationship between the number of bankers and advisors and the growth of bad debts in the portfolio. However, this effect is small and there is no relationship between the level of loan-loss provisions and bankers and advisors. In column 7 and 8, finally, we investigate whether the reforms have an impact on branch-level profitability. There is no evidence that it does. It appears that any increase in sales productivity that bankers and advisors provide comes at a significant cost. In 2007, the bank decided to reduce the number of advisors in small branches because it felt this cost was too high.

VI. Generalized Propensity Score Estimation

The results so far show that the introduction of bankers and advisors had a positive impact on sales productivity, but did not improve profits or other indicators of "quality". These results are based on an IV strategy and within the IV framework the results are very robust. However, the method itself has a weakness in its approach to eliminating endogeneity bias. In particular, IV estimates compare observations with a high expected level of treatment (a high ratio of bankers and advisors to employees) to observations with low expected treatment, but ignore actual treatment levels (Blundell and Costa Dias 2009). As a result, estimates of treatment effects could still be biased by systematic differences between branches that are correlated with both the level of treatment and its expected impact. Arguably, this concern is especially relevant in our context where we construct instruments based on groupings of branches within regions and by size.

In order to further test the robustness of our results, we therefore use a GPS estimator, which compares branches with the same expected treatment but different actual treatment

(Imbens 2000; Hirano and Imbens 2004; Imai and van Dyk 2004). The GPS estimator is similar in spirit to propensity score matching (Rosenbaum and Rubin 1983) and its objective is to eliminate bias that is due to observable differences between treated units from treatment effects estimation. However, unlike propensity score matching, the GPS estimator can be applied to multi-valued or continuous treatments such as the ratio of bankers and advisors per employee. Our implementation of the GPS estimator, which follows (Hirano and Imbens 2004) is sketched out below; Appendix B has a more detailed discussion.

There are two key differences between the IV and GPS estimators. First, GPS estimation involves an explicit before-after comparison. We exploit this to identify separately the performance impact of bankers and that of advisors and delineate two reform periods: (i) the introduction of the bankers in the first four quarters (2003) and (ii) the introduction of the advisors between quarters 11 and 15 (2005/6).[6] In both cases, we generate difference-in-difference estimates that compare branch performance at the beginning of the reform period to branch performance about a year after the reforms were introduced. The second distinction between the IV and the GPS estimates is that the former represent a parametric relationship between treatment and productivity while the latter represent a dose response function: semi-parametric estimates of the impact of treatment on performance at a range of treatment levels.

The first set of GPS estimates in Panel A of Table 4 compare the sales productivity of the bank's branches in quarters 7 to 10 to sales productivity in quarters 1 to 4. To arrive at these estimates, we first estimate the propensity for treatment of each branch as a function of branch characteristics in quarters 1 to 4 on the basis of a fractional logit model (Papke and Wooldridge 1996 , see Table A2.1). Using these estimates, we calculate for each branch the predicted sales

[6] In the IV estimation, we cannot separately include bankers and advisors in the same estimation because eventually, increasing the number of instruments causes multicollinearity in the instrument matrix.

productivity at a range of treatment levels *t*, conditional on the propensity score at *t*. The dose response function represents the average over all branches of the predicted sales productivity at *t* (Appendix B and Hirano and Imbens 2004). The estimate of the treatment effect is calculated as the dose response at *t* minus the dose response at a baseline level, $t = 0$ in our case.

The left-most coefficient in Panel A of Table 4 is a difference-in-difference estimate of the impact of an increase in the number of bankers per employee from zero to 5 percent. The first difference is that between performance in quarters 7 to 10 and performance in quarters 1 to 4, the second difference is that between no treatment and treatment at 5 percent. The other estimates in Table 4 show the dose response function over the range of 10 to 35 percent, which is about the highest observed percentage of bankers per employee in the data.

Despite the fact that our estimates are based on a difference-in-difference specification, the results are broadly consistent with those of the IV approach. The impact of bankers on sales productivity is positive as long as more than 20 percent of branch employees are bankers. At that level, the contribution of bankers to sales productivity is the equivalent of about 2 standard deviations. In Panel B of Table 4 we report estimates of the effect of advisors on sales productivity. In this case, we estimate the propensity for treatment in quarter 15 on the basis of branch characteristics in quarters 9 to 12 (Table A2.1). Using the propensities to estimate treatment effects in Table 4, we find a negative relationship between sales performance and the introduction of advisors.

A. Balancing and Common Support

The literature on propensity score matching has developed tools to evaluate bias reduction characteristics and common support conditions. The toolbox for GPS estimation is still under development, but we build on Hirano and Imbens (2004) and Imai and Van Dyk (2004) to analyze bias reduction and on Flores and Mitnik (2013) to derive common support conditions. Both tests for bias reduction examine the hypothesis that conditional on the GPS, there is no correlation between branch characteristics used to estimate the GPS and the level of treatment. Hirano and Imbens (2004) use a blocking approach to test the hypothesis. Table A2.2 analyzes the differences in branch characteristics between branches in one tertile of treatment levels (the "treated branches") and branches outside of that tertile. Before conditioning on the propensity score (unadjusted difference), 21 out of 51 differences are significant on at least the 5% level of significance. After conditioning on the propensity score (adjusted difference), only 8 differences are significant at that level and both the differences and the t-statistics tend to shrink. Imai and Van Dyk (2004) use a regression-based approach to test for correlation between treatment and branch characteristics. Specifically, they compare t-statistics in a series of regressions of branch characteristics on (i) the treatment variable and (ii), the treatment variable and the propensity for treatment. Figure A2.1 shows that, after we control for the GPS, the distribution of t-statistics on the treatment variable is about normal, consistent with the absence of systematic correlation between treatment and branch characteristics.

We next assess the robustness of the GPS estimates to two sets of common support conditions on the propensity scores. First, we exclude branches with zero treatment from the sample. In preliminary analysis, we found that these branches are quite different from other branches, both in terms of "raw" characteristics and in terms of estimated propensity scores. For the bankers, the results of this exercise are reported in Table 5. Because branches with zero treatment are excluded, the treatment effects are now estimated with 5 percent treatment as a benchmark and the first line reports the estimates from Panel A in Table 4 as a reference. Once we exclude branches with zero treatment, the impact of bankers on sales productivity is positive across the range and there is a concave relationship between bankers per employee and sales productivity, just like in the IV estimates (Panel B in Table 5).

Second, to impose further common support conditions on the remaining, non-zero treatment branches, we adapt a method developed in Flores and Mitnik (2013) for multiple discrete treatments to an environment with continuous treatments. The technical details are again in Appendix B, but the idea is as follows: we first divide the branches into three tertiles by treatment level and calculate the propensity scores for each of the branches at the median level of treatment for their tertile. We then calculate the propensity score for the branches in the other tertiles, also at the median for the first group. If the propensity score for a branch falls outside of the range of scores for the first treatment group, we eliminate the branch from the data. We repeat this procedure taking the other two tertiles as the treatment group, so that our final sample consists of observations that are in the support of all three treatment groups.

With these additional conditions, the estimated treatment effects remain significant and the relationship between bankers per employee and sales productivity remains concave (Panel C of Table 5). Further analysis shows that with the common support conditions imposed, sample balance improves, confirming to the validity of the results.

We also performed balancing tests for the second reform period when the advisor position was introduced. The results (reported in Table A2.3 and Figure A2.2) are less encouraging than those for the bankers: the number of t-statistics that are significant at the 5% level goes down by only two from 13 (unadjusted) to 11 (adjusted) and a number of differences move in the wrong direction. In Figure A2.2, the distribution of t-statistics deviates from the normal distribution even after we control for the GPS. When we impose common support conditions on this set of estimates, the estimated treatment effects change both in sign and size (by an order of magnitude; Table A2.4). However, common support conditions do not improve the balancing properties of the GPS estimates and ultimately we cannot pin down a clear relationship between the introduction of advisors and sales productivity.

The results with regard to the introduction of the bankers are consistent with the results from the IV regression, but the results for the advisors are not. It is possible that the impact of the advisors is simply not robust. To evaluate this, we re-estimated the GMM model in Table 2 with *Bankers / FTE* instead of (*Bankers + Advisors*) */ FTE*. The coefficients in Table A2.5 are smaller than those in Table 2, but they have similar sign and significance, suggesting that the results in Table 2 are driven mostly by bankers.

Focusing on the introduction of bankers only, we use the GPS approach to also analyze the impact of the HRM reforms on profitability and indicators of loan quality in Table 6. The estimates show a negative relationship between the ratio of profit to footing and the presence of bankers. The effect is relatively small (less than one standard deviation), but at odds with the stated goal of the reforms to improve the volume of sales *and* to attract customers that would be more profitable to the bank. The estimates in Table 6 give a mixed picture of loan quality. On the one hand, the introduction of bankers is associated with a reduction in loan loss provisions, i.e. of expected loan losses. At the same time however, the arrival of bankers is associated with an increase in bad debts. Between Table 6 and Table 3, which showed an insignificant relationship between profits, loan quality and the introduction of bankers and advisors, there is no evidence that the reforms had an impact beyond an increase in sales volume.

VII. Conclusion

We exploit an episode of strategic restructuring in a bank – introduction of bankers and advisors with strong individual incentives – to study the impact of modern HRM policies in an emerging market context. The policies created a new functional structure in the bank's branches and high-powered incentives for a subset of branch employees. We find that the reforms have raised the sales productivity in the branches, although the effect appears to be driven by bankers only. The impact of the reforms is larger in large branches, but declines when the ratio of bankers and advisors per employee rises. These results are consistent with the notion that the temptation to free ride is strongest in large branches and that there are limits to the ability or motivation of cashiers to provide support services to bankers and advisors.

We find mixed evidence on the relationship between the HRM reforms and indicators of quality such as profitability, portfolio composition and loan performance. On the one hand, this is good news: despite the fact that the bonus system primarily rewards volume and that the differentiation of incentives creates tensions, loan standards have not deteriorated dramatically. Other papers find a much more negative relationship between the introduction of sales incentives and loan quality (Agarwal and Wang 2009). On the other hand, an important goal of the reforms was to promote the sale of mortgages and sophisticated savings products and to tie high-value customers to the bank.

Research on the relationship between bank performance and foreign acquisition has found that in emerging markets, foreign acquisition improves bank performance (Claessens, Demirguc-Kunt and Huizinga 2001; Bonin, Hasan and Wachtel 2005). The literature has argued that foreign owners improve performance by, among other things, introducing modern management. Our paper provides concrete evidence in support of this argument. At the same time, the results suggest that there are challenges to the implementation of new HRM policies. Our data do not have sufficient detail on the characteristics of branch employees and managers to uncover exactly why the reforms failed to improve quality indicators. However, in a study of a bonus system that was based on a balanced scorecard, Griffith and Neely (2009) point out that inexperienced managers may not be able to balance multiple targets. In the bank, branch managers need to maximize sales, while taking care of quality, the product mix and the fact that differentiation in incentives creates tension in the system. In interviews, we were told that some branch managers found it challenging to control the bankers who essentially sought to run their own franchise within the branch. Hence, it is important that HRM policies are designed taking into account the ability of managers to deal with any tensions that arise in the system. This is especially relevant in emerging markets where managers may not have experience with sophisticated incentive systems or where performance measurement may be more difficult than in advanced economies.

Our IV strategy exploits a unique feature of our data which is that the implementation of HRM reforms is the result of a policy initiative at the level of the bank's headquarters, whereas our data are at the branch-level. In most datasets, the decision to implement new HRM policies is made at the level where it is implemented. This is true of firm-level studies (Ichniowski, Shaw and Prennushi 1997), but also of branch-level studies by (Bartel 2004) and (Bartel, Freeman, Ichniowski and Kleiner 2011), who focus on implementation of policies by branch managers. Although our data is unique in the sense that headquarters provides a policy shock, other research in personnel economics might be able to exploit policy shocks that come from outside the firm. For example, cost reductions in information technology could spur the adoption of organizational reforms and the same could be true of revisions in certain ISO quality standards.

We use GPS estimation to ascertain the robustness of our IV estimates. This is especially useful to test whether our IV results are biased by the fact that there are structural differences between branches with a high, or with a low likelihood of adoption of HRM reforms. Compared to the IV approach, GPS estimation controls more carefully for observable differences between branches and compared to traditional matching estimators GPS estimation allows for multivalued (and multidimensional) treatments. In the context of research on HRM reforms this is a useful property because it should enable researchers to use GPS estimation to evaluate complementarities between HRM policies.

TABLE 1 Summary Statistics and Correlations

Panel A: Branch Staffing and Labor Productivity, by Year and by Size

Year	Branches	Employees	Retail Bankers	SME Bankers	Advisors	ΔLoans / FTE	ΔDeposits / FTE	Profit / FTE
		FTE	% FTE	% FTE	% FTE	Thsnds	Thsnds	Thsnds
Large Branches (20 employees or more)								
2003	49	34.4	6.0%	7.2%				1,404
2004	48	34.8	10.1%	11.5%		2,341	12,265	1,500
2005	45	34.0	10.0%	12.4%	0.4%	4,593	6,831	1,653
2006	47	31.6	11.8%	12.2%	9.1%	9,779	12,676	2,077
2007	43	32.3	12.1%	12.9%	12.0%	10,385	14,674	2,320
Medium-sized Branches (8 to 20 employees)								
2003	78	11.6	4.4%	1.0%				1,221
2004	77	11.6	9.0%	2.8%		1,628	10,732	1,371
2005	72	12.1	9.5%	3.4%	0.7%	4,482	7,620	1,399
2006	63	11.9	10.7%	4.0%	14.1%	8,348	12,969	1,934
2007	64	11.7	10.7%	4.3%	16.0%	12,063	14,277	2,203
Small Branches (7 employees or fewer)								
2003	55	5.4	0.8%					830
2004	54	5.4	1.4%			1,635	9,472	977
2005	63	5.6	1.8%		0.4%	3,356	4,767	1,156
2006	70	5.6	4.5%		8.9%	8,204	10,938	1,564
2007	71	5.2	3.8%		3.8%	9,537	14,552	2,208

Panel B: Correlations, by year (number of observations in *italics*)

	Employees	Retail Bankers	SME Bankers	Advisors	ΔLoans / FTE	ΔDeposits / FTE	Profit / FTE
Retail Bankers	0.402***	1					
SME Bankers	0.618***	0.242***	1				
Advisors	0.030	0.222***	0.062	1			
ΔLoans / FTE	0.022	0.092**	0.099**	0.282***	1		
ΔDeposits / FTE	0.076*	0.041	0.048	0.035	0.470***	1	
Profit / FTE	0.209***	0.248***	0.164***	0.296***	0.227***	0.031	1

Note: FTE is Full Time Equivalent. ΔLoans / FTE and ΔDeposits / FTE are based on loans and deposits outstanding as reported on the balance sheet in local currency at the end of each year. Profit per Employee reflects annual profits per branch (branches with less than 4 quarterly observations in a year are excluded from the calculation of median profit). The reported figures are median values measured in local currency. The correlations in Panel B are based on yearly averages and exclude pre-2005 observations for Advisors and pre-2006 observations for Leadership Academy because Advisors were first introduced in 2005 and the Leadership Academy started in 2006. * significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 2 Sales (ΔFooting/FTE) and Branch Characteristics

	(1) OLS	(2) GMM	(3) GMM
Bankers + Advisors / FTE	0.094	0.374	0.392
	[0.045]**	[0.101]***	[0.116]***
Bankers + Advisors / FTE squared	-0.193	-1.264	-1.463
	[0.122]	[0.437]***	[0.417]***
Bankers + Advisors	0.004	0.010	0.015
	[0.003]	[0.006]*	[0.005]***
Leadership Academy	0.006	0.012	0.011
	[0.006]	[0.007]*	[0.007]
FTE	-0.003	-0.004	-0.004
	[0.001]**	[0.001]***	[0.002]**
FTE Squared	0.000	0.000	0.000
	[0.000]***	[0.000]	[0.000]
Unemployment rate	-0.068	-0.048	-0.052
	[0.041]	[0.037]	[0.038]
Constant	0.083	0.081	0.086
	[0.018]***	[0.017]***	[0.018]***
Observations	3245	3245	3245
Number of Branches	188	188	188
IV/GMM diagnostics (p-values)			
Hansen J test		0.656	0.287
Kleibergen-Paap test for underidentification		0.000	0.000
Difference-in-J test (endogeneity of instrumented variables)		0.004	0.570
First Stage F-statistics			
Bankers + Advisors / FTE		84.69	155.50
Bankers + Advisors / FTE squared		90.35	136.00
Bankers + Advisors		46.95	235.00
FTE			269.10
FTE Squared			47.95

Note: *Footing* is the sum of Loans and Deposits. Δ*Footing /FTE* is the change in footing per employee from quarter t - 1 to quarter t. *Bankers + Advisors* is equal to the number of Retail and SME Bankers and Advisors in a branch. *Leadership Academy* is a dummy that equals 1 when a branch manager has finished the Academy and 0 otherwise. The *unemployment rate* is measured at the level of the administrative district of a branch. In the GMM estimates, instruments for *Bankers + Advisors/FTE* and its square and for *Bankers + Advisors* (and for *FTE* and *FTE squared* in column 3) are constructed from the number of bankers, advisors and employees in other branches in the same region or the same size class (see table 1 for size classes). In particular, the instruments are the average number of employees in the same region, the average number of retail bankers in the same size class, the share of branches in the same region with at least one SME banker, the share of branches in the same size class with at least one advisor , the average number of retail bankers in branches in the same size class in 2003 to 2005 (in columns 3). The instruments also include a categorical variable identifying the phases in the rollout of the program that introduced the banker positions. All models include region x quarter x year fixed effects and city/town dummies for branches located in towns with 50,000 to 100,000 people or cities with more than 100,000 people. Robust standard errors, clustered by branch, in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 3 The Quality of Sales and Branch Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Δ Mortgage /FTE	Δ Mortgage / Loans	Δ Funds / FTE	Δ Funds / Deposits	Δ Profit / FTE	Δ Profit / Footing	Δ Bad Debts / Footing	Δ Provisions / Footing
Bankers + Advisors / FTE	0.058	-0.059	0.274	-0.072	-0.003	0.000	0.006	-0.012
	[0.025]**	[0.160]	[0.087]***	[0.087]	[0.002]	[0.037]	[0.014]	[0.022]
Bankers + Advisors / FTE squared	-0.215	0.012	-0.932	0.249	0.011	-0.035	-0.073	0.046
	[0.100]**	[0.648]	[0.333]***	[0.340]	[0.012]	[0.141]	[0.058]	[0.096]
Bankers + Advisors	0.002	0.002	0.010	-0.003	0.000	0.001	0.002	0.000
	[0.001]*	[0.009]	[0.004]**	[0.004]	[0.000]	[0.002]	[0.001]**	[0.001]
Constant	0.018	0.040	0.013	0.009	0.000	0.004	0.003	0.000
	[0.003]***	[0.012]***	[0.006]**	[0.007]	[0.000]	[0.005]	[0.003]	[0.002]
Observations	2574	2578	2574	2578	3245	3247	3248	3247
Number of Branches	187	187	187	187	188	188	188	188
IV/GMM diagnostics (p-values)								
Hansen J test	0.122	0.774	0.008	0.022	0.459	0.398	0.472	0.432
Kleibergen-Paap test for underidentification	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000

Note: *Bankers + Advisors* is equal to the number of Retail and SME Bankers and Advisors in a branch. *FTE* is the number of employees in a branch. All estimates are done by GMM. *Bankers + Advisors*, Bankers + Advisors / FTE and its square are treated as endogenous. Instruments for Bankers + Advisors / FTE, its square and Bankers + Advisors are constructed on the basis of the presence of bankers, advisors and employees for other branches in the same region or the same size class (see table 1 for size classes). In particular, the instruments are the average number of employees in the same region, the average number of retail bankers in the same size class, the share of branches in the same region with at least one SME banker in the same region, the share of branches with at least one advisor in the same size class. In addition, the instruments include a categorical variable identifying the phases in the rollout of the program that introduced the banker positions. All models include a dummy that is equal to 1 if the branch manager has taken the leadership academy, FTE, FTE squared, the unemployment rate at the district level, region x quarter x year fixed effects and city/town dummies for branches located in towns with 50,000 to 100,000 people or cities with more than 100,000 people. Robust standard errors, clustered by branch, in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 4 Bankers and Advisors and Sales Per Employee - Generalized Propensity Score Estimates

Panel A: Bankers / FTE	5%	10%	15%	20%	25%	30%	35%
ΔFooting / FTE	-0.004	-0.007	-0.001	0.015	0.037	0.060	0.083
	[0.002]*	[0.004]*	[0.004]	[0.004]***	[0.004]***	[0.003]***	[0.002]***
Observations	167	167	167	167	167	167	167

Panel B: Advisors / FTE	5%	10%	15%	20%	25%
ΔFooting / FTE	-0.006	-0.011	-0.019	-0.027	-0.028
	[0.004]	[0.004]***	[0.004]***	[0.004]***	[0.004]***
Observations	178	178	178	178	178

Note: The numbers in this table are estimates of the impact of having a certain share of bankers (advisors) per branch employee (with percentage shares ordered by column) on sales per employee in a branch. The estimates represent difference-in-difference estimates of the dose-response function at various ratios of bankers (advisors) to employees. The first difference is the difference in sales per employee between branches with zero percent bankers per employee and branches with a higher share of bankers per employee and the second difference is between sales per employee in quarters 7 to 10 and sales per employee in quarters 1 to 4 (quarters 17 to 20 minus quarters 9 to 12 for the advisors in panel B). See text and Appendix B for further details. Standard errors are bootstrapped with 1,000 repetitions. * significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 5 Bankers and Sales Per Employee - Generalized Propensity Score Estimates

Bankers / FTE	10%	15%	20%	25%	30%	35%
Panel A: Baseline Estimates (Table 4)						
ΔFooting / FTE	-0.003	0.003	0.019	0.041	0.064	0.087
	[0.003]	[0.004]	[0.004]***	[0.003]***	[0.002]***	[0.001]***
Observations	167	167	167	167	167	167
Panel B: Excluding observations with zero treatment						
ΔFooting / FTE	0.015	0.038	0.058	0.069	0.066	0.049
	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.004]***
Observations	109	109	109	109	109	109
Panel C: Excluding observations with zero treatment and with overlap conditions imposed						
ΔFooting / FTE	0.022	0.040	0.051	0.051	0.042	0.027
	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.003]***
Observations	72	72	72	72	72	72

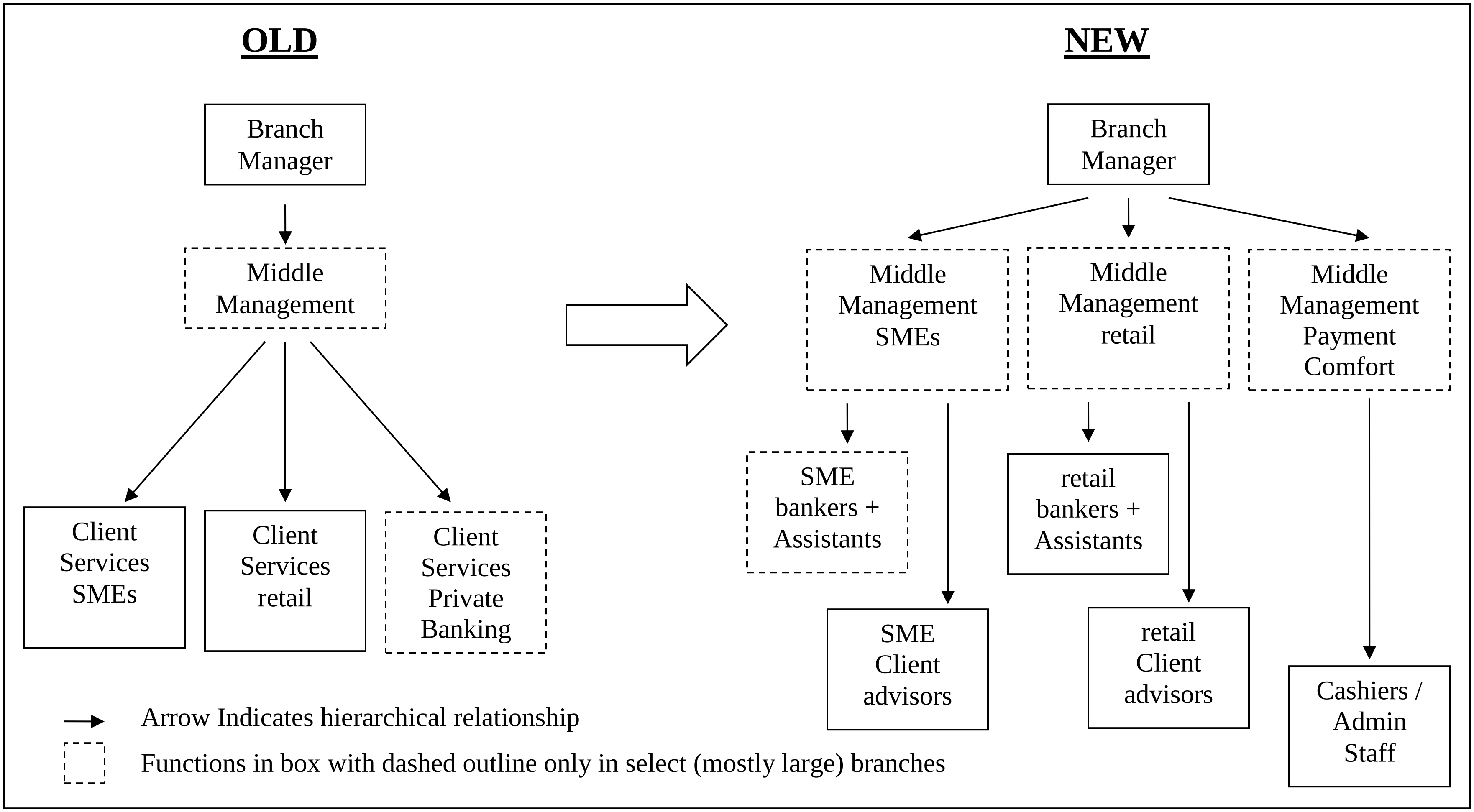
Note: The numbers in this table are estimates of the impact of having a certain share of bankers per branch employee (with percentage shares ordered by column) on sales per employee in a branch. The estimates represent difference-in-difference estimates of the dose-response function at various ratios of bankers to employees. The first difference is the difference in sales per employee between branches with 5 percent bankers per employee and branches with a higher share of bankers per employee and the second difference is between sales per employee in Quarters 7 to 10 and sales per employee in quarters 1 to 4. The first set of estimates is from Table 4. In the second set of estimates, the observations with zero bankers were eliminated. For the third set of estimates, the branches with non-zero treatment were divided into tertiles based on their treatment level. Branches outside of each treatment group are eliminated from the sample if their propensity score at the median treatment in the treatment group is less than the propensity score at the second percentile of propensity scores in the treatment group. See Appendix B for details. Standard errors are bootstrapped with 1,000 repetitions. * significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 6 Bankers and Performance - Generalized Propensity Score Estimates

Bankers / FTE in Quarters 7 to 10	5%	10%	15%	20%	25%	30%	35%
Profit per Employee	0.000	-0.001	-0.002	-0.002	-0.002	-0.002	-0.001
	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***
Provision / Loans	0.000	-0.006	-0.015	-0.024	-0.030	-0.034	-0.035
	[0.000]	[0.000]***	[0.001]***	[0.001]***	[0.002]***	[0.003]***	[0.003]***
Bad Loans / Loans	0.000	0.007	0.014	0.017	0.018	0.016	0.013
	[0.000]	[0.002]***	[0.002]***	[0.003]***	[0.003]***	[0.003]***	[0.002]***
Observations	167	167	167	167	167	167	167

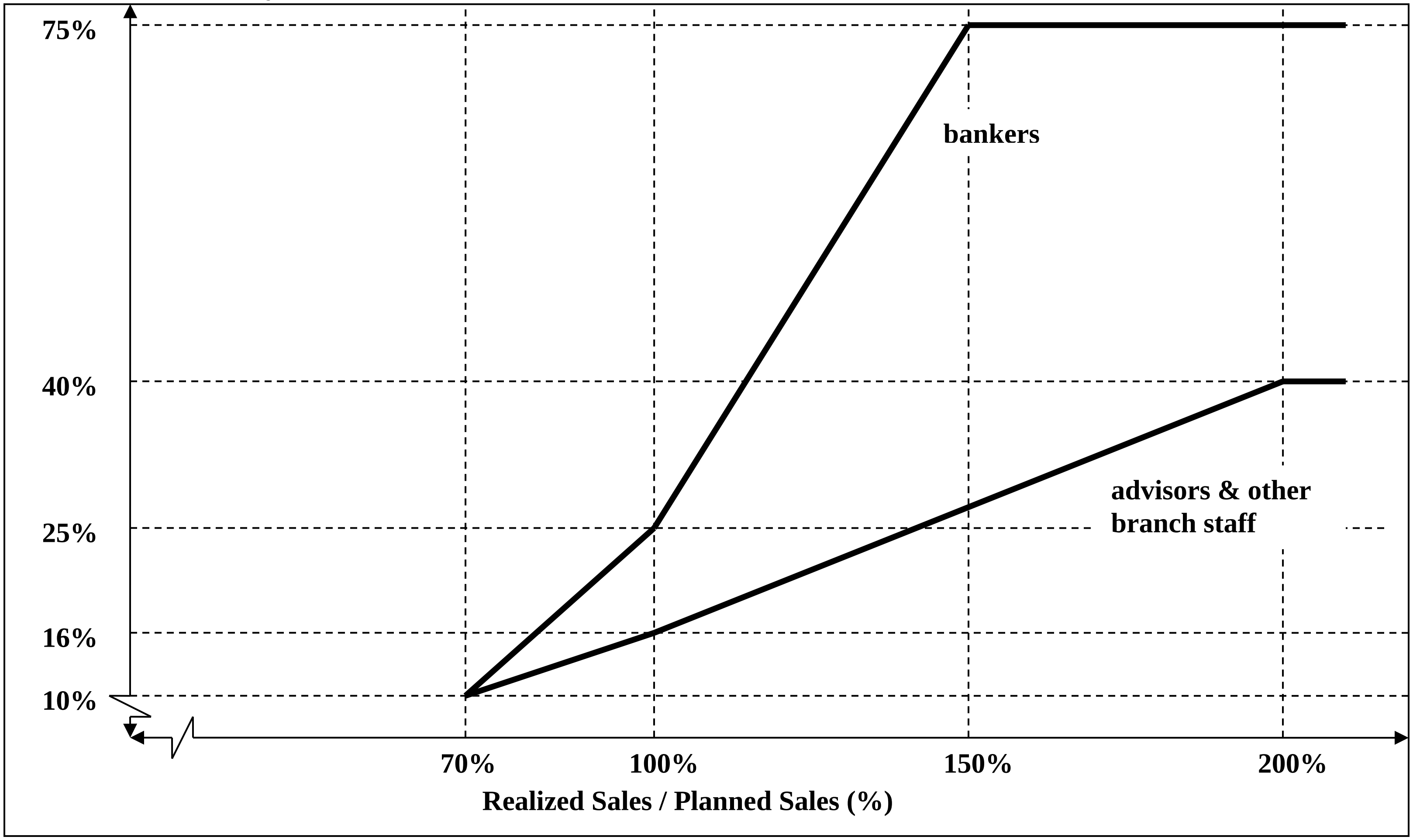
Note: The numbers in this table are estimates of the impact of having a certain share of bankers per branch employee (with percentage shares ordered by column) on profit per employee and on loan loss provisions and bad loans as a percentage of loans in a branch. The estimates reflect the average impact over quarters 7 to 10e. The estimates represent difference-in-difference estimates of the dose-response function at various ratios of bankers to employees. The first difference is the difference in sales per employee between branches with zero percent bankers per employee and branches with a higher share of bankers per employee and the second difference is between performance in quarters 7 to 10 and performance the base period (quarters 1 to 4). See Appendix B for details. Standard errors are bootstrapped with 1,000 repetitions. * significant at 10%; ** significant at 5%; *** significant at 1%

FIGURE 1 Old and New Branch Organizational Models



Note: In the new branch organizational model, the smallest branches have only a branch manager and staff at the Cashier level.

FIGURE 2 Bonus System



Note: bankers' and advisors' final bonus is a 70/30 weighted average of personal and branch performance. All other non-managerial staff receive a bonus based on branch performance.

Appendix A

Estimating treatment effects using the generalized propensity score

The generalized propensity score (GPS) itself and its use in estimating the effect of continuous treatments are developed in Imbens (2000), Hirano and Imbens (2004) and Imai and van Dyk (2004). Our empirical implementation primarily follows Hirano and Imbens (2004).

If we define the treatment (bankers or advisors per employee) as a variable t ò T , the GPS is the density of t conditional on a vector of branch characteristics X. By definition, treatment is independent of X after we condition on the propensity score. The use of the GPS in treatment effects estimation is then justified by the "weak unconfoundedness assumption" (Hirano and Imbens 2004). Specifically, the assumption is that at any t ò T the effect of treatment on the outcome (sales productivity in our case) does not depend on the actual treatment received conditional on X:

$$Y(t) \perp T \mid X \ \forall t \in T \tag{A2.1}$$

If this is true, Hirano and Imbens show, conditioning on the GPS is sufficient to remove any bias associated with differences in branch characteristics X (Imai and Van Dyk (2004), use a slightly different approach to come to essentially the same conclusion).

Hirano and Imbens (2004) suggest a three-step approach to estimate treatment effects using the GPS. The first step is to estimate the propensity for treatment conditional on X. In the second step, the conditional expectation of the outcome Y_i is modeled as a function of the estimated propensity score r_i (t_i, X_i) and actual treatment t_i. The third and final step is to estimate the potential outcomes, Y_i ($r(t, X_i)$, t) at hypothetical treatment levels t on the basis of the estimated relationship between Y, R and T. In and of themselves, the estimates Y_i ($r(t, X_i)$, t) have no causal interpretation because observations with the same propensity score r may not belong to a unique sub-population (in this respect the GPS for multi-valued treatments is different from the GPS for binary treatments, see Imbens 2000). In particular, we cannot compare $E[Y(s) \mid r(s, X) = r]$ to $E[Y(t) \mid r(t, X) = r]$ because $r(s, X) = r$ and $r(t, X) = r$ represent different sub-populations in the data. By the same logic however, we *can* compare $E[Y(s)]$ and $E[Y(t)]$, the population averages of $E[Y(s) \mid r(s, X)]$ and $E[Y(t) \mid r(t, X)]$, because both of these expected values are based on the same population. Consequently, we can obtain a dose-response function $Y(t)$ by calculating the sample average of $Y_i\,(r(t, X_i), t)$ at a series of potential treatments t. Treatment effects are calculated as the difference between $Y(t)$ and the dose-response at a baseline treatment such as $Y(0)$.

A. Implementation

We identify two treatment periods in our data. The first covers the introduction of the bankers in quarters 1 to 4 and the second the introduction of the advisors in quarters 11 to 15 (figure 3). Treatment is measured as the number of bankers or advisors per employee in a branch. This varies between zero and about 0.35 for the bankers and between zero and 0.25 for the advisors.

Because our treatment variable is a fraction, we use a fractional logit model (Papke and Wooldridge 1996) to estimate the propensity score. For the first treatment period, we calculate branch characteristics in X on the basis of data from the first four quarters. Specifically, we estimate:

$$\begin{aligned} t_{ij} = f\,(&FTE_{ij}, \Delta FTE_{ij}, Size_{ij}, Footing_{ij}/FTE_{ij}, \Delta Footing_{ij}/FTE_{ij}, \\ &Unemployment\ rate_{ij}, city_{ij}, region_{j}) \end{aligned} \qquad \text{(A2.2)}$$

All variables were calculated as averages over the first four quarters, except for $city_{ij}$ and $region_{ij}$ (which are constant over time) and ΔFTE_{ij}. The latter variable was calculated as $FTE_{ij4} - FTE_{ij1}$.

In estimating (A2.2), we allow the coefficient on *FTE* to vary between branch size groups to allow for non-linearities in the equation. The results of this first step are reported in Table A2.1.

The second step is to estimate the expected outcome conditional on the propensity score and the level of treatment for each branch. In order to do this, we calculate the estimated propensity for treatment for each branch $\hat{r}_{ij}$ at the observed level of treatment t_{ij} using the estimates in Table A2.1. Following Hirano and Imbens (2004), we then estimate the following quadratic equation:

$$\Delta Y_{ij} = \alpha + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 \hat{r}_{ij}(t_{ij}) + \beta_4 \hat{r}_{ij}^2(t_{ij}) + \beta_5 t_{ij} \times \hat{r}_{ij}(t_{ij}) \qquad \text{(A2.3)}$$

The dependent variable in this equation, ΔY, is the difference between $\Delta Footing_{ij} / FTE_{ij}$ over quarters 1 to 4 and $\Delta Footing_{ij} / FTE_{ij}$ over quarters 7 to 10. As discussed above, the coefficients in this equation do not have a causal interpretation because a given level *r* of *R* does not identify a unique sub-population.

In the third and final step, the coefficients are used to calculate the estimated outcome at each level of treatment and the treatment effects. In particular, we calculate, for each level of treatment and each branch:

$$\Delta \mathrm{Y}_{ij}(t) = \hat{\alpha} + \hat{\beta}_1 \cdot t + \hat{\beta}_2 \cdot t^2 + \hat{\beta}_3 \cdot \hat{r}_{ij}(t) + \hat{\beta}_4 \cdot \hat{r}_{ij}^2(t) + \hat{\beta}_5 \cdot t \cdot \hat{r}_{ij}(t) \qquad \text{(A2.4)}$$

Finally, we use the estimates of outcomes $\Delta Y_{ij}(t)$ for each branch from equation (A2.4) to calculate the estimate of the dose response function at *t*:

$$\hat{E}\left[\Delta Y(t)\right] = \frac{1}{N} \sum_i \Delta Y_{ij}(t) \qquad \text{(A2.5)}$$

The left-hand side of equation (A2.5) represents the expected increase in $\Delta Footing_{ij} / FTE_{ij}$ between quarters 1 to 4 and quarters 7 to 10 at treatment level *t*. Finally, the treatment effect is estimated as:

$$\Delta Y_{t-0} = \hat{E}\left[\Delta Y(t)\right] - \hat{E}\left[\Delta Y_0(0)\right] \tag{A2.6}$$

Where ΔY_{t-0} is a difference-in-difference estimate of the increase in performance associated with an increase in the ratio of bankers to employees from 0 to *t*. The standard errors for the expected performance $\hat{E}\left[\Delta Y(t)\right]$ and the treatment effect need to be corrected for the fact that they are based on estimated propensity scores. Following Hirano and Imbens, we report bootstrapped errors.

The calculation of treatment effects for the introduction of the advisor function follows the same approach as that for the introduction of the banker function. In this case, quarters 9 to 12 are the base period and quarters 17 to 20 are the period in which outcomes are measured. Also, first-step estimation of the propensity score includes the ratio of bankers to FTE. The results of this first-step estimation are displayed in Table A2.1.

B. Balancing and Common Support

The rationale for matching (with binary treatments) and GPS estimation (with multi-valued treatments) is that the propensity score can improve the balance of the sample. However, the GPS is not guaranteed to improve the balance and it is important to assess the extent to which an improvement in balance is indeed achieved. In addition, the estimation of treatment effects requires that there is overlap (or a common support) in the likelihood that branches receive treatment *t* regardless of whether they actually receive *t* or some other treatment *s* (Flores and Mitnik 2013). There are no fully agreed upon methods for the evaluation of balancing properties or the extent of overlap in the context of multi-valued treatments. However, we use methods developed by Hirano and Imbens (2004) and Imai and Van Dyk (2004) to assess balancing and we adapt the method for discrete treatments in Flores and Mitnik (2013) to continuous treatments to evaluate the extent to which there is common support.

Balancing tests. In order to ascertain the balancing properties of the propensity score estimation, we implement two tests. The first one follows Hirano and Imbens (2004) and involves partitioning the observations according to treatment status and propensity scores. The second test follows Imai and Van Dyk (2004) and involves regressing the branch characteristics on the treatment variable with and without the propensity score. In both cases, we test the assumption that, conditional on the propensity score, there are no meaningful differences between branches according to their level of treatment.

Beginning with the first treatment period and following Hirano and Imbens (2004), we divide the branches into three treatment groups of about equal size: a group of branches with no bankers (58 branches), a group of branches with between zero and 0.17 bankers per employee (55) and a group of branches with between 0.17 and 0.36 bankers per employee (54). For each of these groups and each of the branch characteristics we first test whether the branch characteristics in the treatment group are significantly different from the characteristics in the other groups. The t-statistics for this test are reported in Table A2.2, columns 1, 3 and 5. The table shows that there are significant differences between treatment groups, in particular according to size and productivity.

In order to check whether the propensity score improves the balance in the sample, we estimate, for each treatment group, the propensity score at the median of the range of treatments in the group (0 in the case of the first group) and divide the observations in each group into quintiles. We then estimate the propensity score for observations in the other two groups at the same treatment level and assign the observations to the quintiles of the first group according to their propensity score (i.e. if the first group is the treatment group, we calculate the propensity score for branches in the other two groups at $t = 0$ and if a propensity score falls within one of the quintiles of the propensity scores for the treatment group, we group the branch in that quintile).[9]

If branch characteristics are independent of treatment status conditional on the propensity score, the branches within each quintile should be similar to each other regardless of treatment levels. To test whether this is true, we calculate the difference in branch characteristics between branches in the treatment group and those outside of the treatment group for each quintile. We then take observation-weighted averages of each difference across the quintiles and calculate a t-statistic to assess whether the difference is significantly different from zero. The results for the first treatment period are displayed in Table A2.2 and those for the second period are displayed in Table A2.3. For the first treatment period, the propensity score delivers a significant improvement in balance. In the "raw" data, 21 out of 51 t-statistics were higher than 1.96. After adjusting for the propensity score, only 7 are (compare columns 2, 4 and 6 to the "adjusted" t-statistics in columns 1, 3 and 5). For the second treatment period, the number of t-statistics higher than 1.96 drops from 14 before to 11 after taking into account the propensity score (out of a total of 54 this time; see Table A2.3, with the same column comparisons as above).

The risk of relying on t-tests to assess balancing is that a drop in significance of differences may be due to an increase in variance rather than a decrease in actual differences between observations in the treated and non-treated groups. However, comparison of the unadjusted and adjusted differences in Tables A2.2 shows that the adjusted differences tend to be smaller than the unadjusted ones. Unfortunately, this does not apply in Table A2.3.

[9] The results are similar when we use tertiles rather than quintiles.

The second set of balancing tests is based on Imai and Van Dyk (2004). It also seeks to ascertain that, conditional on the propensity score, branch characteristics and treatment are uncorrelated. Instead of partitioning the data into broad treatment groups, Imai and Van Dyk use a regression-based approach that evaluates differences in the covariates along their support in the data. Following their example, we run two series of regressions. The first series are regressions of the branch characteristics used in the estimation of the GPS on the treatment variable (using linear regressions for continuous characteristics and logit regressions for binary variables[10]). The second series are the same regressions, but including the GPS in addition to the treatment variable. If the GPS properly balances the sample, its inclusion should render the treatment insignificant. Figure A2.1 presents the results for the first treatment period. The figure plots the quantiles of the t-statistics on the treatment variable (*Bankers / FTE*) against the quantiles of the normal distribution. The GPS clearly improves the balance and brings the distribution of t-statistics much closer to the normal distribution.

Figure A2.2 repeats the exercise for the second treatment period, with *Advisors / FTE* as the treatment variable. In this case, including the GPS in the regressions slightly narrows the distribution of t-statistics but much less so than in the first reform period. This reinforces the conclusion from Table A3 that the GPS does little to improve sample balance in the second reform period.

Common Support. In binary treatment models, the efficacy of matching estimators in reducing bias is contingent on the presence of a common support for propensity scores between treated and non-treated individuals (Heckman, Ichimura and Todd 1997). Definition of common support is more complicated in the case of multi-valued treatments because the propensity score for each

[10] Using linear regressions for binary variables produces similar results.

individual has to be evaluated at multiple treatment values against the propensity score of individuals receiving a particular level of treatment (Flores and Mitnik 2013).

Our approach to evaluating the role of a common support is to assess the robustness of our estimates to the imposition of more stringent common support conditions. We do this in two steps. First, we exclude observations that receive zero treatment (i.e. zero bankers or zero advisors). Analysis of the propensity scores at $t = 0$ revealed that branches with zero treatment are extremely likely to receive no treatment, but that the propensity of other branches to receive zero treatment is very low. Conversely, the propensities of zero treatment branches to receive non-zero treatment are lower than of almost any of the branches that receive non-zero treatment.

Second, after excluding the branches with zero treatment, we adapt the method developed in Flores and Mitnik (2013) to impose commonality of support among branches with non-zero treatment. Flores and Mitnik (2013) have data with discrete treatments. To ensure that there is overlap in the support between observations, they calculate the propensity score for each observation at each treatment $t \in T$. At each t, they calculate a cutoff value q_t, which is defined as the second percentile of the distribution of propensity scores among the individuals receiving t. Subsequently, they calculate the propensity for other individuals to receive treatment t. Individuals are excluded from the sample if their propensity score at any t is lower than q_t.

We adapt this procedure to continuous treatments by dividing the branches with non-zero treatment into tertiles on the basis of their treatment. Starting with the first tertile, we calculate the propensity score for treatment at the median of this treatment group, m_1. We then determine the propensity score at the second percentile of the treatment group q_1 and for the second and third tertiles we use a similar procedure to calculate m_2, q_2 and m_3, q_3. Finally, we exclude all branches that are not in the first tertile, but have a propensity score at m_1 that is smaller than q_1

and we similarly exclude branches that have no common support with the branches in the second and third tertiles.

We assess the extent to which our estimates are robust to the imposition of stricter conditions for common support in Table 5 (main text, for the bankers) and Table A2.4 (for the advisors). Because we exclude branches with non-zero treatment, we use t = 5% as our new baseline treatment (hence, the estimated treatment effects in Tables 5 and A2.4 represent $\Delta\mathrm{Y}_{t-5} = \hat{E}\left[\Delta\mathrm{Y}(t)\right] - \hat{E}\left[\Delta Y(5)\right]$). For ease of reference, we have included the comparable estimates from Table 4 in Tables 5 and A2.4. The results in Table 5 suggest that the estimates for the bankers in Panel A of Table 4 largely hold up when we impose a common support. The estimated effects remain positive and they are of similar magnitude although they are somewhat higher at low levels of treatment and reveal diminishing returns at higher levels of treatment. Further analysis of these results reveals that the restriction of the sample also improves the balance (both before and after controlling for the GPS).

The results in Table A2.4 by contrast are more mixed. After imposing common support conditions, the sign of the estimated treatment effects changes. However, it is not clear that the estimates with the common support conditions are more reliable than the estimates without those conditions: the sample balance does not improve after imposing these conditions and conditional on the GPS it may even get worse. Hence, the GPS estimates do not give us solid evidence that the introduction of the advisors improved branch performance, but they confirm that the introduction of the banker positions was good for branch performance.

TABLE A2.1 Propensity score estimation (fractional logit)

	(1) **Bankers / FTE**	(2) **Advisors / FTE**
FTE	-0.008 [0.003]**	-0.015 [0.004]***
FTE x Size (8 to 20 employees)	0.080 [0.021]***	0.012 [0.013]
FTE x Size (20 employees or more)	1.511 [0.608]**	0.383 [0.088]***
ΔFTE	0.211 [0.107]**	0.110 [0.434]
Bankers and Advisors / FTE		-1.511 [0.664]**
Bankers and Advisors / FTE x Size (8 to 20 employees)		0.624 [0.985]
Bankers and Advisors / FTE x Size (20 employees or more)		-0.236 [1.691]
Size (8 to 20 employees)	-1.651 [0.297]***	-0.083 [0.252]
Size (20 employees or more)	-12.546 [4.293]***	-2.853 [0.633]***
Footing / FTE	1.088 [0.317]***	0.325 [0.223]
ΔFooting / FTE	0.116 [0.639]	-1.854 [1.424]
Unemployment rate	0.405 [1.829]	-1.565 [1.789]
Population 0 to 50,000	0.241 [0.142]*	-0.131 [0.095]
Population 50,000 to 100,000	0.158 [0.132]	-0.161 [0.098]
Region 1	-0.114 [0.255]	0.293 [0.145]**
Region 2	0.150 [0.262]	-0.082 [0.179]
Region 3	-0.064 [0.178]	0.347 [0.198]*
Region 4	-0.060 [0.218]	0.028 [0.145]
Region 5	0.053 [0.175]	0.114 [0.173]
Region 6	0.269 [0.205]	
Region 7	0.008 [0.201]	-1.050 [0.332]***
Observations	167	175

Note: The dependent variables are Bankers / FTE in quarter 7 (column 1) and Advisors / FTE in quarter 15 (columns 2). Robust standard errors in brackets. * significantly different from 0 at 10%; ** significant at 5%; *** significant at 1%

TABLE A2.2 Balancing tests (table 4, Panel A)

treatment group	Bankers / FTE = 0				Bankers / FTE = 0.061 to 0.167				Bankers / FTE = 0.170 to 0.353			
	(1)		(2)		(3)		(4)		(5)		(6)	
	Unadjusted difference	t-test	Adjusted difference	t-test	Unadjusted difference	t-test	Adjusted difference	t-test	Unadjusted difference	t-test	Adjusted difference	t-test
FTE	15.942	7.53	1.010	2.11	4.518	1.84	8.001	2.20	-21.077	-11.20	-9.415	-2.63
ΔFTE	0.040	0.71	-0.036	-1.85	0.009	0.15	0.100	1.20	-0.050	-0.88	-0.062	-0.53
Size (8 to 20 employees)	0.387	5.16	0.042	0.24	-0.586	-8.62	-0.268	-2.87	0.190	2.35	0.310	3.10
Size (20 employees or more)	-0.782	-17.10	-0.042	-0.24	0.338	4.66	-0.028	-0.39	0.469	6.87	0.031	0.84
Footing	0.279	7.90	0.097	1.33	-0.053	-1.28	-0.018	-0.31	-0.235	-6.18	-0.095	-1.49
ΔFooting	-0.004	-0.39	0.014	1.47	-0.024	-2.53	-0.031	-1.74	0.028	2.96	0.036	2.03
Unemployment rate	-0.017	-2.28	-0.001	-0.05	0.005	0.60	0.014	1.32	0.013	1.70	0.003	0.31
Population 50,000 to 100,000	0.202	3.81	0.000		0.007	0.12	0.091	1.01	-0.216	-4.02	-0.025	-0.28
Population > 100,000	0.324	4.81	0.125	1.24	-0.176	-2.46	-0.280	-2.69	-0.158	-2.19	0.029	0.24
Region 1	0.043	0.79	0.167	1.00	-0.075	-1.34	0.008	0.09	0.030	0.54	0.071	0.76
Region 2	-0.012	-0.24	-0.042	-0.24	0.034	0.71	0.050	0.75	-0.023	-0.46	-0.058	-0.73
Region 3	-0.123	-2.18	0.042	0.29	0.079	1.36	0.079	1.03	0.048	0.83	-0.015	-0.27
Region 4	-0.007	-0.12	-0.042	-0.24	-0.021	-0.34	-0.002	-0.02	0.029	0.46	0.075	0.79
Region 5	-0.063	-1.23	0.083	0.44	0.007	0.13	-0.038	-0.55	0.059	1.11	-0.012	-0.17
Region 6	0.166	3.16	0.042	0.29	-0.029	-0.54	-0.002	-0.03	-0.143	-2.64	-0.158	-1.55
Region 7	0.041	0.86	-0.083	-0.81	-0.020	-0.41	-0.154	-2.12	-0.023	-0.46	0.066	0.80
Region 8	-0.045	-0.83	-0.167	-1.20	0.025	0.45	0.059	0.83	0.022	0.39	0.031	0.38

Note: This table reports improvements in the balance of the sample (used in Table 4) after controlling for the Generalized Propensity Score. In order to implement the balancing tests, the sample was split in tertiles on the basis of treatment levels. The "unadjusted difference" represents the difference in averages of the covariates between branches in a given tertile of treatments (the "treated" branches) and branches outside of the tertile. To calculate the "adjusted difference", the generalized propensity scores of all observations in a given tertile of treatments were split into quintiles. Subsequently, based on their propensity scores, the non-treated branches are are assigned to the quintiles. The adjusted difference represents the observation-weighted difference between branches in each of the quintiles (see the text of Appendix B for further details).

TABLE A2.3 Balancing tests (table 4, Panel B)

treatment group	Advisors / FTE = 0.000 to 0.125				Advisors / FTE = 0.129 to 0.167				Advisors / FTE = 0.176 to 0.250			
	(1)		(2)		(3)		(4)		(5)		(6)	
	Unadjusted difference	t-test	Adjusted difference	t-test	Unadjusted difference	t-test	Adjusted difference	t-test	Unadjusted difference	t-test	Adjusted difference	t-test
FTE	-8.838	-4.63	-13.173	-4.20	2.438	1.18	1.890	0.77	7.304	3.48	9.706	3.78
ΔFTE	0.031	1.74	-0.009	-0.33	-0.026	-1.40	-0.030	-1.45	-0.015	-0.76	-0.007	-0.28
Retail Bankers /FTE	-0.019	-1.27	-0.092	-4.27	-0.013	-0.82	0.007	0.42	0.038	2.39	0.085	4.98
Size (8 to 20 employees)	0.378	5.32	0.151	1.33	0.098	1.25	0.133	1.56	-0.522	-7.22	-0.196	-2.94
Size (20 employees or more)	-0.073	-0.98	0.286	2.61	-0.087	-1.15	-0.131	-1.53	0.178	2.27	-0.149	-1.91
Footing	-0.005	-0.09	-0.224	-2.70	-0.009	5.32	0.151	1.33	0.047	0.87	0.195	3.03
Δfooting	-0.010	-1.75	-0.001	-0.12	0.005	-0.98	0.286	2.61	0.005	0.86	0.007	0.89
Unemployment rate	0.004	0.52	0.019	1.95	-0.011	1.18	1.890	0.77	0.003	0.48	-0.004	-0.47
Population 0 to 50,000	-0.132	-2.56	-0.001	-0.01	0.016	-0.82	0.007	0.42	0.124	2.24	0.098	1.47
Population 50,000 to 100,000	0.049	0.68	-0.219	-1.86	0.048	-1.40	-0.030	-1.45	-0.051	-0.66	0.100	1.03
Region 1	0.100	1.97	-0.070	-1.06	0.007	-0.18	0.010	0.17	-0.134	-2.47	-0.095	-1.35
Region 2	-0.041	-0.89	-0.124	-1.46	0.016	0.81	0.003	0.43	0.022	0.44	0.047	0.72
Region 3	-0.111	-1.98	0.058	0.71	0.007	-1.57	-0.021	-2.91	0.108	1.78	-0.034	-0.49
Region 4	0.226	4.17	0.170	2.02	-0.196	0.29	-0.007	-0.11	-0.059	-0.97	-0.017	-0.23
Region 5	-0.002	-0.05	0.100	1.12	0.024	0.65	0.120	1.57	-0.035	-0.61	-0.006	-0.08
Region 6	0.074	1.17	-0.128	-1.17	0.040	0.14	-0.016	-0.26	-0.065	-0.94	0.060	0.68
Region 7	0.000		0.000		0.000	0.34	0.061	1.21	0.000		0.000	
Region 8	-0.246	-5.25	-0.007	-0.12	0.100	0.12	0.045	0.66	0.164	3.11	0.044	1.03

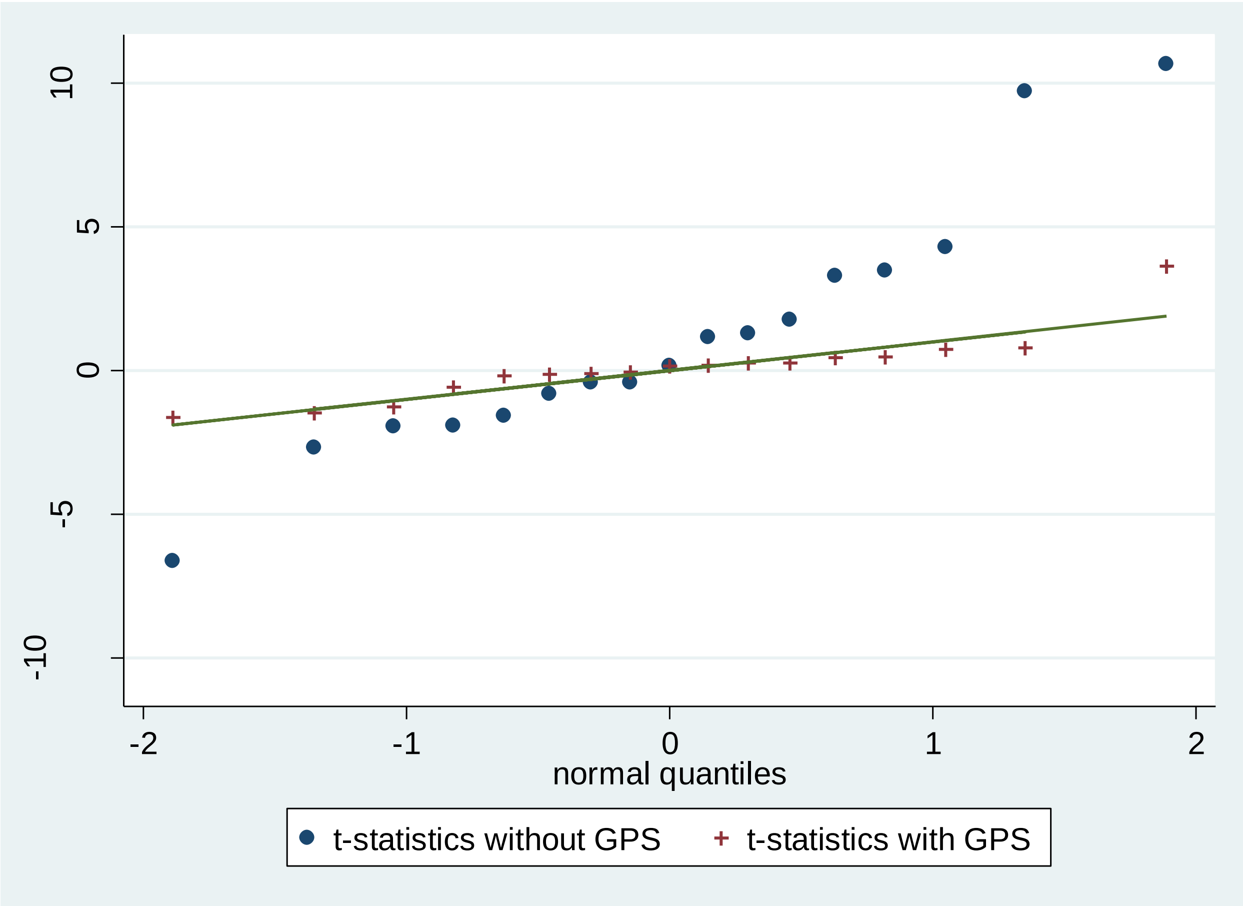
Note: This table reports improvements in the balance of the sample (used in Table 5) after controlling for the Generalized Propensity Score. In order to implement the balancing tests, the sample was split in tertiles on the basis of treatment levels. The "unadjusted difference" represents the difference in averages of the covariates between branches in a given tertile of treatments (the "treated" branches) and branches outside of the tertile. To calculate the "adjusted difference", the generalized propensity scores of all observations in a given tertile of treatments were split into quintiles. Subsequently, based on their propensity scores, the non-treated branches are are assigned to the quintiles. The adjusted difference represents the observation-weighted difference between branches in each of the quintiles (see the text of Appendix B for further details).

TABLE A2.4 Impact of Advisors on Sales Per Employee - Generalized Propensity Score Estimates

Advisors / FTE	10%	15%	20%	25%
Baseline Estimates (Table 4, panel B)				
	-0.006	-0.014	-0.021	-0.023
	[0.002]***	[0.002]***	[0.002]***	[0.002]***
Observations	178	178	178	178
Excluding observations with zero treatment				
	-0.021	-0.035	-0.039	-0.032
	[0.001]***	[0.001]***	[0.001]***	[0.001]***
Observations	148	148	148	148
Excluding observations with zero treatment and with overlap conditions imposed				
	0.210	0.253	0.236	0.214
	[0.027]***	[0.027]***	[0.027]***	[0.027]***
Observations	70	70	70	70

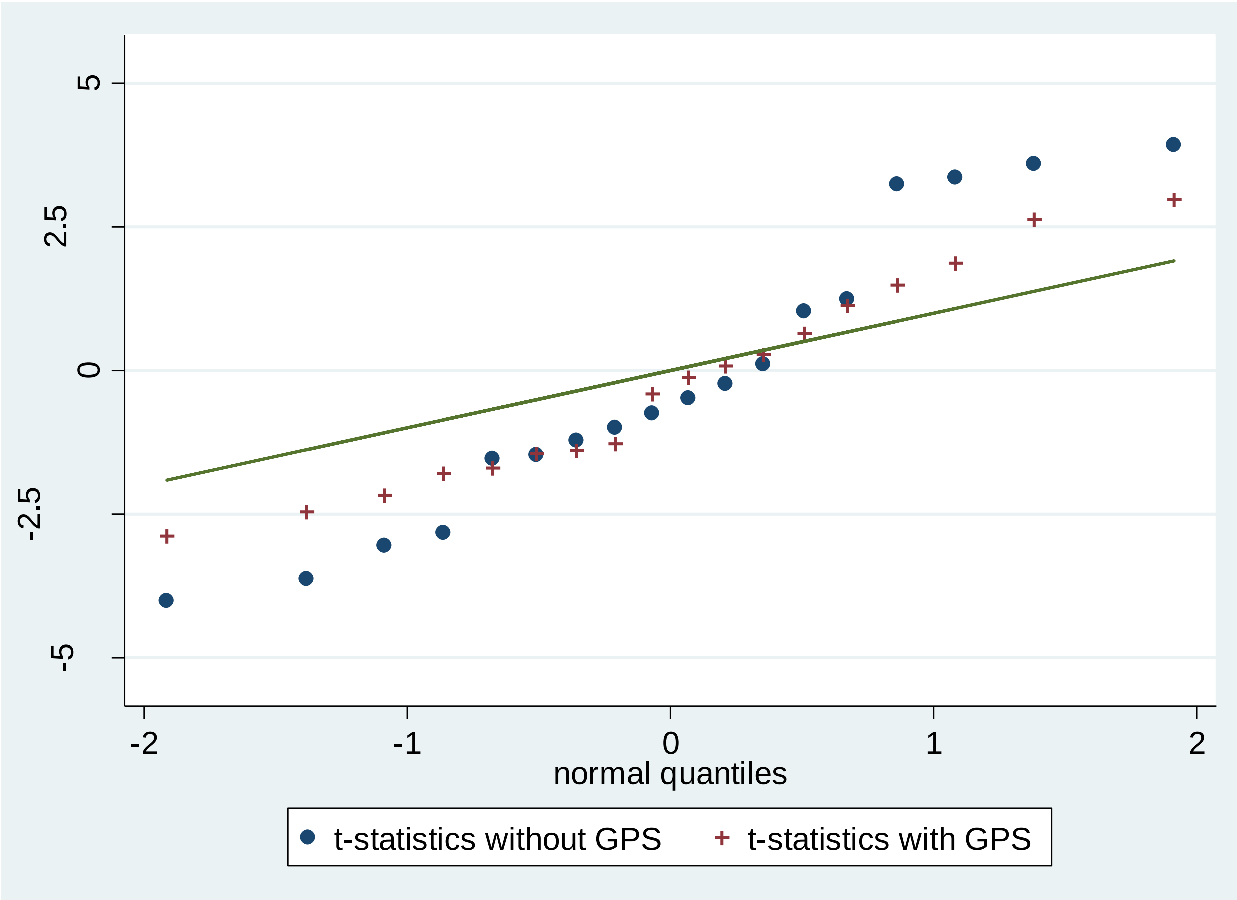
Note: The numbers in this table are estimates of the impact of having a certain share of bankers per branch employee (with percentage shares ordered by column) on sales per employee in a branch. The estimates represent difference-in-difference estimates of the dose-response function at various ratios of bankers to employees. The first difference is the difference in sales per employee between branches with 5 percent bankers per employee and branches with a higher share of bankers per employee and the second difference is between sales per employee in Quarters 17 to 20 and sales per employee in quarters 9 to 12. The first set of estimates is from Table 5. In the second set of estimates, the observations with zero bankers were eliminated. For the third set of estimates, the branches with non-zero treatment were divided into tertiles based on their treatment level. Branches outside of each treatment group are eliminated from the sample if their propensity score at the median treatment in the treatment group is less than the propensity score at the second percentile of propensity scores in the treatment group. See Appendix B for details. Standard errors are bootstrapped with 1,000 repetitions. * significant at 10%; ** significant at 5%; *** significant at 1%

FIGURE A2.1 First Treatment Period



Note: this figure plots the results of regressions of branch characteristics used in the estimation of the generalized propensity score (see table A1) on the treatment variable (Bankers / FTE) in quarter 7. The t-statistics "without GPS" are t-statistics on Bankers/FTE from regressions that include only Bankers/FTE and a constant. The t-statistics "with GPS" are t-statistics on Bankers/FTE from regressions that include the GPS as well as Bankers/FTE and a constant.

FIGURE A2.2 Second Treatment Period



Note: this figure plots the results of regressions of branch characteristics used in the estimation of the generalized propensity score (see table A1) on the treatment variable (Advisors / FTE) in quarter 13. The t-statistics "without GPS" are t-statistics on Bankers/FTE from regressions that include only Bankers/FTE and a constant. The t-statistics "with GPS" are t-statistics on Bankers/FTE from regressions that include the GPS as well as Bankers/FTE and a constant.

References

Agarwal, Sumit, and Hefei Wang, "Perverse Incentives at the Banks? Evidence from a Natural Experiment," FRB of Chicago Working Paper No. 2009-08, 2009.

Alchian, Armen A., and Harold Demsetz, "Production, Information Costs, and Economic Organization," *American Economic Review*, 62 (1972), 777-795.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14 (1999), 57-67.

Athey, Susan, and Scott Stern, "An Empirical Framework for Testing Theories About Complimentarity in Organizational Design," NBER Working Paper No. W6600, 1998.

---, "The Impact of Information Technology on Emergency Health Care Outcomes," *Rand Journal of Economics*, 33 (2002), 399-432.

Baker, George, "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources*, 37 (2002), 728-751.

Bartel, Ann P., "Human Resource Management and Organizational Performance: Evidence from Retail Banking," *Industrial & Labor Relations Review*, 57 (2004), 181.

Bartel, Ann P., Richard B. Freeman, Casey Ichniowski, and Morris M. Kleiner, "Can a Workplace Have an Attitude Problem? Workplace Effects on Employee Attitudes and Organizational Performance," *Labour Economics*, 18 (2011), 411-423.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman, "Enhanced Routines for Instrumental Variables/GMM Estimation and Testing," Center for Economic Reform and Transformation, Heriot-Watt University Working Paper No. 2007/06, 2007.

Berglof, Erik, and Patrick Bolton, "The Great Divide and Beyond: Financial Architecture in Transition," *Journal of Economic Perspectives*, 16 (2002), 77-100.

Besanko, David, Pierre Regibeau, and Katherine E. Rockett, "A Multi-Task Principal-Agent Approach to Organizational Form," *Journal Of Industrial Economics*, 53 (2005), 437-467.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David J. McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *Quarterly Journal Of Economics*, 128 (2013), 1-51.

Blundell, Richard, and Monica Costa Dias, "Alternative Approaches to Evaluation in Empirical Microeconomics," *Journal of Human Resources*, 44 (2009), 565-640.

Bonin, John P., Iftekhar Hasan, and Paul Wachtel, "Bank Performance, Efficiency and Ownership in Transition Countries," *Journal of Banking & Finance*, 29 (2005), 31-53.

Buch, Claudia M., "Opening up for Foreign Banks: How Central and Eastern Europe Can Benefit," *Economics of Transition*, 5 (1997), 339-366.

Cappelli, Peter, and David Neumark, "Do "High-Performance" Work Practices Improve Establishment-Level Outcomes?," *Industrial & Labor Relations Review*, 54 (2001), 737-775.

Claessens, Stijn, Asli Demirguc-Kunt, and Harry Huizinga, "How Does Foreign Entry Affect Domestic Banking Markets?," *Journal of Banking & Finance*, 25 (2001), 891-911.

Commander, Simon J., Rupert Harrison, and Naercio Menezes-Filho, "Ict and Productivity in Developing Countries: New Firm-Level Evidence from Brazil and India," *Review Of Economics And Statistics*, 93 (2011), 528-541.

Corts, Kenneth S., "Teams Versus Individual Accountability: Solving Multitask Problems through Job Design," *Rand Journal of Economics*, 38 (2007), 467-479.

Flores, Carlos A., and Oscar A. Mitnik, "Comparing Treatments across Labor Markets: An Assessment of Nonexperimental Multiple-Treatment Strategies," *Review Of Economics And Statistics*, (2013).

Freeman, Richard B., and Morris M. Kleiner, "The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production," *Industrial Relations: A Journal of Economy and Society*, 44 (2005), 307-330.

Griffith, Rachel, and Andrew Neely, "Performance Pay and Managerial Experience in Multitask Teams: Evidence from within a Firm," *Journal Of Labor Economics*, 27 (2009), 49-82.

Hausman, Jerry A., "Valuing the Effect of Regulation on New Services in Telecommunications," *Brookings Papers on Economic Activity*, (1997), 1-54.

Hausman, Jerry A., and William E. Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica*, 49 (1981), 1377-1398.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64 (1997), 605-654.

Hirano, Keisuke, and Guido W. Imbens, "The Propensity Score with Continuous Treatments," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, A. Gelman, and X.-L. Meng, eds. (New York: Wiley, 2004).

Holmstrom, Bengt, and Paul Milgrom, "Multitask Principal Agent Analyses - Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law Economics & Organization*, 7 (1991), 24-52.

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

---, "Insider Econometrics: A Roadmap to Estimating Models of Organizational Performance," in *Handbook of Organizational Economics*, Robert Gibbons, and John Roberts, eds. (Princeton, NJ: Princeton University Press, 2012).

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87 (1997), 291-313.

Imai, Kosuke, and David. A. van Dyk, "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score," *Journal Of The American Statistical Association*, 99 (2004), 854-866.

Imbens, Guido W., "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika*, 87 (2000), 706-710.

Laffont, Jean-Jacques, and Jean-Charles Rochet, "Collusion in Organizations," *Scandinavian Journal of Economics*, 99 (1997), 485-495.

Lazear, Edward P., "The Power of Incentives," *American Economic Review*, 90 (2000), 410-414.

Murphy, Kevin J., "Performance Standards in Incentive Contracts," *Journal Of Accounting & Economics*, 30 (2000), 245-278.

Papke, Leslie E., and Jeffrey M. Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 11 (1996), 619-632.

Rosenbaum, Paul R., and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70 (1983), 41-55.

Sealey, Calvin W., and James T. Lindley, "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions," *Journal Of Finance*, 32 (1977), 1251--1266.

Shirley, Mary M., and Lixin Colin Xu, "Empirical Effects of Performance Contracts: Evidence from China," *Journal of Law Economics & Organization*, 17 (2001), 168-200.

Tirole, Jean, "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics & Organization*, 2 (1986), 181-214.

Wageman, Ruth, "Interdependence and Group Effectiveness," *Administrative Science Quarterly*, 40 (1995), 145-180.

Weitzman, Martin L., "The Ratchet Principle and Performance Incentives," *Bell Journal Of Economics*, 11 (1980), 302-308.