## Online Appendices

## Online Appendix A - Rosenbaum Sensitivity Test

The primary assumption that underlies matching methods is the conditional independence assumption. In other words, the validity of a matching approach depends on the assumption that the treatment is a random assignment conditional on a set of observed variables, which can be denoted as:

$$
\left(\mathrm{Y}_{0}, \mathrm{Y}_{1}\right) \perp \mathrm{D} \mid \mathrm{X} .
$$

The conditional independence assumption will be satisfied if $X$ includes all of the variables that simultaneously influence both the franchising decision $D$ and the outcome variables (survival and growth rate). Clearly, this is a strong assumption, which might not hold even with a very rich set of observed variables $X$. If there are unobserved variables that affect both the franchising decision and the outcome variables at the same time, the estimated average treatment effects suffer from "hidden bias" problem. Thus, it is important to check the sensitivity of the estimated results with respect to the potential "hidden bias."

Rosenbaum (2002) proposed a bounding approach to measure how the estimated average treatment effects change when an unobserved confounding factor is present in the treatment assignment. Assume two entrepreneurs with the exactly the same observed characteristics, i.e. $X_{i}=X_{j}$, then if the conditional independence assumption holds, we would expect the probability of starting a franchised business to be equal for the two entrepreneurs, i.e. $p\left(X_{i}\right)=p\left(X_{j}\right)$. However, if there is an unobserved confounding factor, then let us assume the following bounds on the odds-ratio that either of the two entrepreneurs will start a franchised business as:

$$
\frac{1}{e^{\gamma}} \leq \frac{p\left(X_{i}\right)\left(1-p\left(X_{j}\right)\right)}{p\left(X_{j}\right)\left(1-p\left(X_{i}\right)\right)} \leq e^{\gamma} .
$$

If $e^{\gamma}=1$, then we assume that there are no confounding factors and the two entrepreneurs who have the same observed characteristics have the same probability of starting a franchised business. If $e^{\gamma}=2$, then we assume that the unobserved confounding factor can cause the odds-ratio of starting a franchised business with the two entrepreneurs who have the same observed characteristics to differ by as much as a factor of two. In this sense, $e^{\gamma}$ is the bound for how much the unobserved confounding factor affects the treatment assignment. If the estimated results are still significant, given a large value of $e^{\gamma}$, then we conclude that the estimated average treatment effects using the matching approach is relatively insensitive to the "hidden bias." Otherwise, the estimated results are sensitive to the "hidden bias."

When the outcome variable is binary, we use the Mantel and Haenszel (MH) test statistic implemented by Becker and Caliendo (2007).

## Online Appendix B - Instrumental Variable Approach and Control Function Approach for the Binary Dependent Variable Case

In our application, as the endogenous variable is binary, we employ both the IV approach and CF approach with caution. Following Wooldridge (2010), for simplicity, we assume a linear probability model for binary outcomes and a linear model for continuous outcomes:

$$
\mathrm{y}_{1}=\mathrm{z}_{1} \delta_{1}+\alpha_{1} \mathrm{y}_{2}+\mathrm{u}_{1}
$$

where $y_{1}$ are the outcomes of interest, either the 1 -year, 2 -year, or 3 -year business survival indicators or the growth rate in employment, $y_{2}$ is the possibly endogenous binary variable, here the franchise business indicator variable, and $\boldsymbol{z}_{\mathbf{1}}$ is a vector of exogenous explanatory variables. To estimate the effect of the franchising decision on business survival, we first estimate a probit model for $\mathrm{y}_{2}$ :

$$
\mathrm{y}_{2}=\mathrm{I}\left[\mathbf{z} \delta_{2}+\mathrm{e}_{2} \geq 0\right],
$$

where $\mathbf{z}$ is a vector including both $\mathrm{z}_{1}$ and the instrumental variables, in this case, including the state-level proportions of franchised businesses in the same industry and in related industries, and the state-level franchisor ownership rate among establishments of franchised chains in the same industry. From this estimation, we obtain the predicted probabilities $\Phi\left(\mathbf{z} \hat{\mathbf{\delta}}_{2}\right)$, where $\hat{\delta}_{2}$ is the probit estimator.

In the IV approach, we use the predicted probabilities $\Phi\left(\mathbf{z} \hat{\delta}_{2}\right)$ as the instrumental variable and 2SLS to get consistent estimates.

In the CF approach, we use the predicted probabilities $\Phi\left(\mathbf{z} \hat{\delta}_{2}\right)$ to calculate the inverse Mills ratio (IMR): $\lambda(\cdot)=\phi(\cdot) / \Phi(\cdot)$ and the "generalized residual:"

$$
\widehat{\mathrm{gr}}_{2} \equiv \mathrm{y}_{2} \lambda\left(\mathbf{z} \widehat{\delta}_{2}\right)-\left(1-\mathrm{y}_{2}\right) \lambda\left(-\mathbf{z} \widehat{\delta}_{2}\right) .
$$

In the second step, we use the "generalized residual" as an additional variable in the following regression:

$$
\mathrm{y}_{1}=\mathrm{z}_{1} \delta_{1}+\alpha_{1} \mathrm{y}_{2}+\gamma_{1} \widehat{\mathrm{gr}}_{2}+e_{1} .
$$

The consistency of the control function estimator hinges on the following three assumptions: (1) $D\left(y_{2} \mid \boldsymbol{z}\right)$ is correctly specified; (2) $\left(\mathrm{u}_{1}, \mathrm{e}_{2}\right)$ is independent of $z_{1}$; and (3) $E\left(u_{1} \mid e_{2}\right)=\rho_{1} e_{2}$, i.e. a linear relationship between the two error terms exists.

## Online Appendix C - Intermediate Results in IV and CF Approaches

Table C1. Summary Statistics for Instrumental Variables

|  | Franchised <br> Businesses |  | Independent <br> Businesses |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Mean | S.D. | Mean | S.D. |  |
| State-Level Franchise Proportion in the Same Industry | 0.37 | 0.30 | 0.18 | 0.21 |  |
| State-Level Franchise Proportion in the Related Industries | 0.26 | 0.18 | 0.20 | 0.19 |  |
| State-Level Franchisor Ownership Rate among Establishments of | 0.18 | 0.17 | 0.20 | 0.23 |  |
| Franchise Chains in the Same Industry | 1,150 |  | 7,850 |  |  |
| Number of Observations |  |  |  |  |  |

Table C2. The First Step in IV and CF Approach

| VARIABLES | Probit Model <br> I(Franchised Business) |
| :--- | :---: |
| State-Level Franchise Proportion in the Same Industry | $1.328^{* * *}$ |
| State-Level Franchise Proportion in Related Industries | $(0.164)$ |
|  | 0.478 |
| State-Level Franchisor Ownership Rate among Establishments | $(0.383)$ |
| of Franchise Chains in the Same Industry | $-0.428^{* * *}$ |
| Business Owner Characteristics | $(0.128)$ |
| Business Characteristics | Yes |
| Market Level Characteristics | Yes |
| FE - Industry | Yes |
| FE - State | Yes |
| FE - Survey Year | Yes |
| Observations | Yes |

Table C3. The First Stage regression in IV Approach

|  | I(Franchised Business) |
| :--- | :---: |
| PARIABLES | $1.126^{* * *}$ |
|  | $(0.069)$ |
| Business Owner Characteristics | Yes |
| Business Characteristics | Yes |
| Market-Level Characteristics | Yes |
| FE - Industry | Yes |
| FE - State | Yes |
| FE - Survey Year | Yes |
| Observations | 9,000 |
| R-Squared | 0.187 |
| Cragg-Donald Wald F statistic | 285 |

Note: The Stock-Yogo weak identification test critical value for this regression is 16.38 . As the estimated CraggDonald Wald F statistic is much greater than this critical value, we reject the null hypothesis of weak identification at the standard $10 \%$ level. The reported results are for the first stage in the unconditional survival analysis. The results of conditional survival analysis and the growth analysis are similar.

## Online Appendix D - Supplementary Tables

Table D1: Summary Statistics

| Means | Businesses Started in |  |  | Businesses Started in <br> $\mathbf{2 0 0 0 7}$ |  | Business Started in |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2012 |  |  |  |  |  |  |  |  |


| Partnership | 0.20 | 0.22 | 0.20 | 0.28 | 0.33 | 0.28 | 0.28 | 0.45 | 0.26 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Corporation | 0.59 | 0.66 | 0.58 | 0.52 | 0.55 | 0.52 | 0.33 | 0.33 | 0.33 |
| Operated from Home | ND | ND | ND | ND | ND | ND | 0.10 | 0.06 | 0.11 |
| Family Owned | 0.41 | 0.51 | 0.40 | 0.37 | 0.53 | 0.34 | 0.24 | 0.42 | 0.22 |
| Initial Year Employment | 1.25 | 2.03 | 1.12 | 1.51 | 2.74 | 1.31 | 2.00 | 3.59 | 1.82 |
| Financing - Bank Loan | 0.31 | 0.46 | 0.29 | 0.28 | 0.42 | 0.25 | 0.12 | 0.27 | 0.11 |
| Financing - Government Guaranteed Loan | 0.06 | 0.13 | 0.05 | 0.05 | 0.10 | 0.04 | 0.03 | 0.06 | 0.02 |
| Financing - Government Loan | 0.04 | 0.06 | 0.04 | 0.03 | 0.06 | 0.02 | ND | ND | ND |
| Financing - Savings | 0.70 | 0.67 | 0.71 | 0.72 | 0.70 | 0.72 | 0.71 | 0.65 | 0.71 |
| Financing - Other Personal Assets | 0.23 | 0.22 | 0.23 | 0.16 | 0.17 | 0.16 | 0.12 | 0.15 | 0.11 |
| Financing - Credit Card | 0.19 | 0.13 | 0.20 | 0.22 | 0.17 | 0.23 | 0.16 | 0.09 | 0.17 |
| Financing - None Needed | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| Financing - Outside Investor | 0.06 | 0.07 | 0.06 | 0.33 | 0.39 | 0.32 | ND | ND | ND |
| Financing - Home Equity Loan | N/A | N/A | N/A | 0.22 | 0.29 | 0.20 | 0.07 | 0.07 | 0.07 |
| Financing - Loan from Family/Friends | N/A | N/A | N/A | 0.07 | 0.06 | 0.07 | 0.07 | 0.06 | 0.07 |
| Financing - Other Sources of Capital | N/A | N/A | N/A | 0.04 | 0.04 | 0.04 | 0.05 | 0.07 | 0.05 |
| Financing - Grants | N/A | N/A | N/A | ND | ND | ND | ND | ND | ND |
| Financing - Venture Capital | N/A | N/A | N/A | ND | ND | ND | ND | ND | ND |
| Financing - Do Not Know | N/A | N/A | N/A | ND | ND | ND | 0.07 | 0.08 | 0.07 |
| Market-Level Variables (Counties) |  |  |  |  |  |  |  |  |  |
| Population | 699,8 | 685,0 | 702,1 | 750,2 | 637,6 | 768,3 | 963,6 | 858,2 | 975,8 |
| pulatio | 02 | 51 | 38 | 54 | 17 | 31 | 57 | 63 | 24 |
| Personal Income (Current \$) | $\begin{gathered} 31,25 \\ 4 \end{gathered}$ | $\begin{gathered} 31,41 \\ 0 \end{gathered}$ | $\begin{gathered} 31,22 \\ 9 \end{gathered}$ | $\begin{gathered} 39,74 \\ 6 \end{gathered}$ | $\begin{gathered} 39,50 \\ 1 \end{gathered}$ | $\begin{gathered} 39,78 \\ 5 \end{gathered}$ | $\begin{gathered} 40,79 \\ 8 \end{gathered}$ | $\begin{gathered} 41,00 \\ 2 \end{gathered}$ | $\begin{gathered} 40,77 \\ 5 \end{gathered}$ |
| Males | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 | 0.49 |
| Caucasian | 0.83 | 0.82 | 0.83 | 0.80 | 0.80 | 0.80 | 0.78 | 0.78 | 0.77 |
| African American | 0.12 | 0.13 | 0.12 | 0.13 | 0.14 | 0.13 | 0.14 | 0.14 | 0.14 |
| Other Race | 0.05 | 0.05 | 0.05 | 0.07 | 0.06 | 0.07 | 0.08 | 0.08 | 0.08 |
| Age: 14 and Younger | 0.21 | 0.21 | 0.21 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
| Age: 15-64 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Age: 65 and Older | 0.13 | 0.12 | 0.13 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 0.13 |
| At Least High School Degree | 0.87 | 0.87 | 0.87 | 0.86 | 0.88 | 0.86 | 0.86 | 0.87 | 0.85 |
| Unemployed | 0.06 | 0.06 | 0.06 | 0.05 | 0.05 | 0.05 | 0.10 | 0.09 | 0.10 |
| Geographic Area (Square Miles) | 1,203 | 1,178 | 1,208 | 1,309 | 1,329 | 1,306 | 1,317 | 1,070 | 1,345 |
| MSA Central County | 0.66 | 0.71 | 0.65 | 0.69 | 0.74 | 0.69 | 0.74 | 0.74 | 0.74 |
| MSA Outside County | 0.08 | 0.08 | 0.08 | 0.09 | 0.08 | 0.09 | 0.07 | 0.09 | 0.07 |
| Non-MSA County | 0.26 | 0.21 | 0.27 | 0.22 | 0.18 | 0.23 | 0.19 | 0.17 | 0.19 |
| Observations | 5,240 | 720 | 4,520 | 1,690 | 230 | 1,460 | 2,100 | 200 | 1,900 |

Notes: Variables unavailable in the SBO surveys are denoted by N/A and cells that failed the disclosure criteria are denoted as ND. Source: U.S. Census Bureau SBO 2002, SBO 2007, SBO 2012, and the LBD.

Table D2: Determinants of the Franchise Decision

| VARIABLES | I(Franchised Business) |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2002 |  |  | 2007 |  |  | 2012 |  |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Male | $0.032 * * *$ | 0.027*** | 0.022** | 0.052*** | 0.05*** | 0.041** | $0.045^{* * *}$ | 0.043*** | 0.036** |
|  | (0.010) | (0.010) | (0.010) | (0.019) | (0.019) | (0.019) | (0.014) | (0.014) | (0.014) |
| African American | ND | ND | ND | ND | ND | ND | 0.004 | 0.014 | 0.023 |
|  | ND | ND | ND | ND | ND | ND | (0.037) | (0.036) | (0.035) |
| Asian | -0.028** | -0.017 | -0.014 | -0.055** | -0.032 | -0.026 | -0.008 | -0.002 | -0.001 |
|  | (0.014) | (0.014) | (0.014) | (0.023) | (0.022) | (0.022) | (0.018) | (0.018) | (0.019) |
| Other Race | ND | ND | ND | ND | ND | ND | -0.005 | 0.007 | 0.011 |
|  | ND | ND | ND | ND | ND | ND | (0.03) | (0.029) | (0.029) |
| Age: 35-44 | 0.027** | 0.024** | 0.021* | 0.042* | 0.029 | 0.018 | 0.017 | 0.018 | 0.017 |
|  | (0.012) | (0.012) | (0.012) | (0.022) | (0.022) | (0.022) | (0.018) | (0.018) | (0.018) |
| Age: 45-54 | 0.030** | 0.027** | 0.023* | 0.036 | 0.018 | $0.000$ |  |  |  |
|  | (0.013) | (0.012) | (0.012) | $(0.023)$ | (0.023) | $(0.023)$ | $(0.018)$ | $(0.018)$ | $(0.018)$ |
| Age: 55-64 | $-0.002$ |  |  | 0.052* | 0.043 | 0.019 | 0.027 | 0.021 | 0.017 |
|  | $(0.015)$ | $(0.015)$ | $(0.016)$ | (0.03) | (0.03) | (0.03) | (0.021) | (0.021) | (0.021) |
| Age: 65 and Older | ND | ND | ND | ND | ND | ND | ND | ND | ND |
|  | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| Technical or Vocational School | -0.002 | -0.003 | -0.005 | ND | ND | ND | ND | ND | ND |
|  | (0.016) | (0.016) | (0.016) | ND | ND | ND | ND | ND | ND |
| Some College | 0.047*** | 0.046*** | $0.044^{* * *}$ | 0.025 | 0.023 | 0.021 | 0.010 | 0.007 | 0.000 |
|  | (0.013) | $(0.013)$ | $(0.012)$ | (0.026) | (0.025) | (0.025) | (0.017) | (0.018) | (0.018) |
| Associate Degree | 0.080*** | 0.080*** | 0.075*** | 0.021 | 0.018 | 0.012 | 0.038 | 0.035 | 0.020 |
|  | (0.02) | (0.02) | (0.02) | (0.031) | (0.031) | (0.031) | (0.03) | (0.03) | (0.029) |
| Bachelor's Degree | 0.099*** |  |  |  |  |  | 0.089*** | 0.077*** | $0.066^{* * *}$ |
|  | $(0.013)$ | $(0.013)$ | $(0.013)$ | $(0.024)$ | $(0.024)$ | $(0.024)$ | (0.019) | $(0.019)$ | $(0.019)$ |
| Postgraduate Degree | $0.088 * * *$ | $0.077 * * *$ | $0.071 * * *$ |  |  | $0.074^{* *}$ | $0.074 * * *$ | $0.063 * *$ | $0.050^{*}$ |
|  | (0.019) | $(0.019)$ | (0.019) | $(0.034)$ | $(0.033)$ | $(0.033)$ | $(0.026)$ | (0.026) | $(0.026)$ |
| Financing - Bank Loan |  | 0.078*** | 0.073*** |  | 0.082*** | 0.075*** |  | 0.124*** | 0.104*** |
|  |  | (0.011) | (0.011) |  | (0.022) | (0.022) |  | (0.027) | (0.027) |
| Financing - Government Guaranteed Loan |  | 0.154*** | 0.149*** |  | 0.164*** | 0.145*** |  | 0.138** | 0.129** |
|  |  | (0.026) | (0.026) |  | (0.053) | (0.053) |  | (0.063) | (0.063) |
| Financing - Government Loan |  | 0.002 | 0.000 |  | 0.121 | 0.120 |  | ND | ND |
|  |  | (0.029) | (0.029) |  | (0.079) | (0.077) |  | ND | ND |
| Financing - Savings |  | -0.020* | -0.022* |  | 0.007 | 0.005 |  | 0.002 | 0.001 |
|  |  |  | 8 |  |  |  |  |  |  |



Notes: Coefficients included in the regressions that cannot be disclosed due to U.S. Census Bureau restrictions are denoted by ND (Not Disclosed). Owner-level default categories are Female, Caucasian, Age: Younger than 35, and Less than High School and High School. ${ }^{* * *}$, **, and * indicate statistical significance at the $1 \%$, $5 \%$, and $10 \%$ levels, respectively. Robust standard errors are shown in parentheses. Market-level controls include measures on Population, Personal Income, Males, Caucasian, African American, Other Race, Age: 14 and Younger, Age: 15-64, Age: 65 and Older, At Least High School Degree, Unemployed, Geographic Area, MSA Central County, MSA Outside County, and Non-MSA County.

Table D3: The Effect of Franchising on Business Survival with Interactions between Franchise and Legal Form of Organizations


|  | $(0.000)$ | $(0.001)$ | $(0.001)$ | $(0.000)$ | $(0.001)$ | $(0.001)$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Financing - Bank Loan | $0.029^{* * *}$ | $0.044^{* * *}$ | $0.047^{* * *}$ | $0.030^{* * *}$ | $0.044^{* * *}$ | $0.047^{* * *}$ |
|  | $(0.006)$ | $(0.010)$ | $(0.013)$ | $(0.006)$ | $(0.010)$ | $(0.013)$ |
| Financing - Gov. Guaranteed | $0.032^{* * *}$ | $0.048^{* * *}$ | $0.076^{* * *}$ | $0.033^{* * *}$ | $0.049^{* * *}$ | $0.076^{* * *}$ |
|  | $(0.010)$ | $(0.017)$ | $(0.022)$ | $(0.010)$ | $(0.017)$ | $(0.022)$ |
| Financing - Gov. Loan | -0.005 | -0.015 | -0.041 | -0.004 | -0.013 | -0.041 |
|  | $(0.015)$ | $(0.023)$ | $(0.029)$ | $(0.015)$ | $(0.023)$ | $(0.029)$ |
| Financing - Other Sources | 0.011 | $0.031^{* *}$ | 0.025 | 0.011 | $0.031^{* *}$ | 0.025 |
|  | $(0.009)$ | $(0.012)$ | $(0.018)$ | $(0.009)$ | $(0.012)$ | $(0.018)$ |
| Financing - Savings | 0.005 | 0.015 | 0.014 | 0.005 | 0.015 | 0.013 |
|  | $(0.007)$ | $(0.010)$ | $(0.013)$ | $(0.007)$ | $(0.010)$ | $(0.013)$ |
| Financing - Other Personal |  |  |  |  |  | -0.001 |
| Assets | $0.017^{* *}$ | -0.001 | -0.002 | $0.017^{* *}$ | -0.002 |  |
|  | $(0.007)$ | $(0.011)$ | $(0.014)$ | $(0.007)$ | $(0.011)$ | $(0.014)$ |
| Financing - Credit Card | -0.010 | $-0.030^{* * *}$ | $-0.058^{* * *}$ | -0.009 | $-0.030^{* * *}$ | $-0.058^{* * *}$ |
|  | $(0.008)$ | $(0.011)$ | $(0.014)$ | $(0.008)$ | $(0.011)$ | $(0.014)$ |
| Financing - None Needed | 0.005 | -0.027 | 0.010 | 0.004 | -0.027 | 0.008 |
|  | $(0.020)$ | $(0.027)$ | $(0.034)$ | $(0.020)$ | $(0.027)$ | $(0.034)$ |
| County-Level Market Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| FE - Industry (NAICS4) | Yes | Yes | Yes | Yes | Yes | Yes |
| FE - State | Yes | Yes | Yes | Yes | Yes | Yes |
| FE - Survey Year | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9,000 | 9,000 | 9,000 | 9,000 | 9,000 |  |
| R-squared | 0.000 | 0.073 | 0.047 | 0.061 | 0.073 |  |

Notes: Owner age and education level are also used as controls but not reported for brevity. ${ }^{* * *},{ }^{* *}$, and $*$ indicate statistical significance at the $1 \%, 5 \%$, and $10 \%$ levels, respectively. Robust standard errors are reported in parentheses.

