Who Distributed More Paycheck Protection Program (PPP) Loans? A Nonparametric Approach to Determine the Characteristics of Banks that Influenced Their Ability to Distribute PPP Loans

Noara Razzak *

August 31, 2023

Abstract

The Paycheck Protection Program (PPP) was an emergency measure taken during Covid-19 pandemic to support small businesses that faced mandated business closures. The businesses were supposed to use PPP loans to help fund payroll costs, including benefits, and also use these loans to pay for mortgage interest, rent, utilities, worker protection costs related to the Covid-19 pandemic. Using Federal Deposit Insurance Corporation (FDIC) Call Report data from June 2020, I measure how much PPP loans/assets were given out by FDIC registered banks. I look at commercial and industrial loan commitments to large businesses, commercial and industrial loans issued for small businesses, unused commercial and industrial loan commitments, core deposits and liquid assets as predictors for the amount of PPP loans given out per bank. I use nonparametric multivariate Kernel regressions and semiparametric smooth coefficient Kernel regressions to understand what institutional features lead to a higher amount of PPP loans/asset for every bank.

Keywords: Paycheck Protection Program

*John E. Walker Department of Economics, School of Business, Clemson University, Clemson, South Carolina, 29634-1309, USA; email nrazzak@clemson.edu. I am grateful to Paul W. Wilson for his insightful comments and advice.

1 Introduction

The Coronavirus Aid, Relief, and Economic Security Act (CARES) Act, passed during the depths of the Covid-19 pandemic enabled the Small Business Administration (SBA), with support from the Department of the Treasury, to implement the Paycheck Protection Program (PPP). Since businesses and schools were forced to close during the second of week of March 2020, the program was designed to provide small businesses with funds to pay up to 8 weeks of payroll costs including benefits to workers. These funds were also allocated so that small businesses could pay interest on mortgages, rent, and utilities. The employment effects of the PPP has been studied extensively, primarily by Autor et al. (2022), Barraza et al. (2020), Granja et al. (2020) and Faulkender et al. (2021). Most of these studies conclude that PPP loans boosted employment at eligible firms at the peak of the pandemic in the middle of 2020 and find statistically and economically significant effects from the program on unemployment. The effects of the PPP in minority and under-served communities have been studied by Howell et al. (2021) and Lester and Wilson (2023). Lester and Wilson (2023) find that majority Black and Hispanic neighborhoods received disproportionately fewer PPP loans than majority White and Asian neighborhoods, similarly, Howell et al. (2021) find that Black-owned businesses were more likely to obtain their PPP loan from a fintech lender than a traditional bank.

The distributional effects of the PPP program have been examined by Bartik et al. (2020), Balyuk et al. (2020), Li and Strahan (2020) and Wang and Kang (2023). Bartik et al. (2020) use a novel firm-level survey data that contains information on banks' relationships with their primary commercial and industrial (C&I) loan customers as well as publicly available data from the SBA to measure heterogeneity in the process of PPP distribution and to assess whether banks targeted loans to high-impact firms. They find that banks did target loans to their most valuable pre-existing customers. Similarly, Balyuk et al. (2020) find that larger borrowers enjoy earlier PPP access, an effect that is more pronounced from borrowers doing business with big banks. Using spatial regressions and in-depth interviews with small businesses, banks, government agencies, and nonprofit organizations, Wang and Kang (2023) finds that the PPP have reasonably succeeded in reaching their small business targets. Nevertheless, they also conclude that communities with higher shares of pandemic-vulnerable businesses or higher levels of socioeconomic vulnerability did not receive a proportional amount of PPP loans. Li and Strahan (2020) take a different approach and mostly delve into the characteristics of the banks that distributed these loans and among other things look at whether C&I loan commitments to large businesses, C&I loan commitments to small businesses, core deposits, liquid assets among other features influenced the amount of PPP loans distributed by these banks.

The relationship between small businesses and their lending banks have been studied extensively by Avery and Samolyk (2004), Amiram and Rabetti (2020), Nguyen (2019) and Nguyen and Barth (2020). Allen and Whitledge (2022) find that during the first week of available PPP loans, community banks issued nearly three times as many loans as large non-community banks. In this paper, I use one of the models advanced by Li and Strahan (2020) to test whether previous C&I loans distributed to small businesses played a role in determining how effective these banks were in distributing PPP loans. However, instead of using a linear model to explain the relationship between the amount of PPP loans distributed in June 2020 and C& I loans distributed by these banks, their core deposits and liquid assets over the four quarters of 2019, I use multivariate Kernel regression, semiparametric smooth coefficient Kernel regression and "Wild" bootstrapping to draw inferences regarding the explanatory variables. The rest of the paper is divided into the following parts. In section 2, I describe the existing economic literature and provide details of the econometric model used. In section 3, I describe the data. Section 4 contains estimation and results and section 5 concludes the paper.

2 The Model

I use multivariate Kernel regressions to test whether the amount of PPP loans distributed by banks as reported by FDIC Call Reports from June 2020 is influenced by previous quarters' C&I loans to large and small businesses, core deposits, liquid assets and unused C&I loan commitments. The predictors were chosen primarily to test out Li and Strahan (2020)'s parametric linear model. Nevertheless, previous literature have used these variables to predict future C&I lending of banks. Avery and Berger (1991) reason that although high loan commitments increase a bank's risk exposure, they also help the bank to ration and sort out riskier borrowers, thereby making the banks more robust in the long run. Kashyap et al. (2002) study how a bank's lending activities and deposit-taking activities are inextricably linked and makes banks "a very special type of financial institution". Berlin and Mester (1999). Harvey and Spong (2001) study the importance of core deposits in a bank's ability to maintain loan commitments in response to exogenous shocks to credit risks and increased competition among other financial service providers. Finally, Kim and Sohn (2017) find that the growth rate of net loans and unused loan commitments, is positively associated with the level of bank liquidity.

Below, I provide a brief discussion of the local linear Kernel estimator as shown in García-Portugués (2023). The local mean Kernel regression estimator was introduced by Èlizbar Nadaraya (1964). Later, local linear and other higher order estimators were introduced as detailed in Wand and Jones (1995), Fan and Gijbels (1996) and Li and Racine (2007). In Kernel regression, the population object to be estimated is the function $m : \mathbf{R}^p \to \mathbf{R}$ and is given by:

$$m(x) := \mathbf{E}[Y|X = x] = \int y f_{y|X=x}(y) dy \tag{2.1}$$

where $\mathbf{X} = (X_1, ..., X_p)'$ is the vector of predictors. We can also rewrite m as:

$$m(x) = \frac{\int yf(\mathbf{x}, y)dy}{f_{\mathbf{x}}(\mathbf{x})} \tag{2.2}$$

Here, f is the joint density of $(\mathbf{X}, Y)$ and $f_{\mathbf{x}}$ is the marginal probability density function (pdf) of $\mathbf{X}$. Given a sample $(\mathbf{X}_i, Y_i)$, we can estimate f and $f_{\mathbf{X}}$ by the kernel density estimators

$$\hat{f}(\mathbf{x}, y; \mathbf{H}, h) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)K_h(y - Y_i), \tag{2.3}$$

$$\hat{f}(\mathbf{x}; \mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) \tag{2.4}$$

Using 2.3 and 2.4 and plugging into 2.2, we get the Nadaraya-Watson estimator for multivariate predictors:

$$\hat{m}(\mathbf{x}; 0, \mathbf{H}) := \sum_{i=1}^{n} \frac{K_{\mathbf{H}}(\mathbf{x} - \mathbf{X_i})}{\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X_j})} Y_i = \sum_{i=1}^{n} W_i^0(\mathbf{x})Y_i \tag{2.5}$$

Here,

$$W_i^0(\mathbf{x}) := \frac{K_{\mathbf{H}}(\mathbf{x} - \mathbf{X_i})}{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X_i})}$$

and $\mathbf{H}$ is the diagonal bandwidth such that,

$$\mathbf{H} = diag(h_1^2, ..., h_p^2) = diag(\mathbf{h^2})$$

The estimator used by the "np" package in R by Hayfield and Racine (2008) uses the following estimator derived from 2.5

$$\hat{m}(\mathbf{x}; 0, \mathbf{h}) := \sum_{i=1}^{n} W_i^0(\mathbf{x})Y_i$$

Here,

$$W_i^0(\mathbf{x}) := \frac{K_{\mathbf{h}}(\mathbf{x}-\mathbf{X_i})}{\sum_{i=1}^{n} K_{\mathbf{h}}(\mathbf{x}-\mathbf{X_i})}$$

and

$$K_{\mathbf{h}}(\mathbf{x}-\mathbf{X_i}) = K_{h_1}(x_1 - X_{i1} \times ..^{p}. \times K_{h_p}(x_p - X_{ip}).$$

Therefore, the Nadaraya-Watson estimate is a weighted average of $Y_1, ..., Y_n$ by means of the weights $\{W_i^0((x)\}_{i=1}^{n}$, in other words, the Nadaraya-Watson estimator is a local mean of $Y_1, ..., Y_n$ about $\mathbf{X} = \mathbf{x}$. The local linear estimator involves the first order Taylor expansion of the function defined in 2.2.

$$m(\mathbf{X_i}) \approx m(\mathbf{x}) + Dm(\mathbf{x})^{'}(\mathbf{X_i} - \mathbf{x})$$

The estimate for $m(\mathbf{x})$ is obtained from the solution of the weighted least squares problem given below

$$\begin{aligned}\hat{\beta_{\mathbf{h}}} &= \arg\min_{\beta \in \mathbf{R}^{p+1}} (\mathbf{Y} - \mathbf{X}\beta)^{'}\mathbf{W}(\mathbf{Y} - \mathbf{X}\beta) \\ &= (\mathbf{X}^{'}\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}^{'}\mathbf{W}\mathbf{Y})\end{aligned}$$

where,

$$\mathbf{W} := diag(K_{\mathbf{x}}(\mathbf{X_1} - \mathbf{x}), ..., K_{\mathbf{x}}(\mathbf{X_n} - \mathbf{x}))$$

Hence, the estimator function $\hat{m}(\mathbf{x}; 1, h)$ can be expressed as

$$\begin{aligned}\hat{m}(\mathbf{x}; 1, h) &:= \hat{\beta}_{\mathbf{h},0} \\ &= \mathbf{e}_1^{'}(\mathbf{X}^{'}\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}^{'}\mathbf{W}\mathbf{Y}) \\ &= \sum_{i=1}^{n} W_i^1(\mathbf{x}) Y_i\end{aligned} \tag{2.6}$$

where

$$W_i^1(\mathbf{x}) := \mathbf{e}_1^{'}(\mathbf{X}^{'}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{W}\mathbf{e_i}$$

García-Portugués (2023) states that the local linear estimator is a "weighted linear combination of the responses". This linear combination is not a weighted mean, as in the case of the Nadaraya-Watson estimator, since the weights $W_i^1(\mathbf{x})$ can be negative despite $\sum_{i=1}^{n} W_i^1(\mathbf{x}) = 1$.

As noted by García-Portugués (2023), the following are the only assumptions required for the asymptotic analysis of the local polynomial estimator:

- m is twice continuously differentiable.
- $\sigma^2 := Var[Y|X = x]$, the conditional variance of Y given X is continuous and positive.
- The marginal pdf of X is continuous and bounded away from zero.
- The kernel K is a symmetric and bounded pdf with finite second moment and is square integrable.
- $h = h_n$ is a deterministic sequence of bandwidths such that, when $n \to \infty$, $h \to 0$ and $nh \to \infty$.

  Due to the fewer assumptions required by both the local linear and the local constant estimators when compared with parametric estimators, the degree of model misspecification is substantially muted. As I demonstrate in the Estimation and Results section, the linear parametric models are also rejected by the data.

3 The Data

I use FDIC Call Report data from the four quarters of 2019 and the second quarter of 2020, which has information on the amount of PPP loans banks committed to small businesses in the spring and early summer of 2020. The variable RCONLG27 in the June 2020 Call Report contains information on outstanding PPP loan amounts. The variable RCFD2170 (or RCON2170) contains information on total assets for the banks, the variable RCON1763 contains C&I loan commitments for large businesses, the variables RCON 5571 and RCON5573 contain C&I loan commitments to small businesses less than $250,000 and the variable RCON5575 contains C&I commitments to small businesses between $250,000 and $1,000,000. The variable RCON2215 contains core deposits amount for each bank. Liquid assets are calculated by adding up the variables RCFD0081(or RCON0081), RCFD0071 (or RCON0071), RCFD1773 (or RCON1773) and RCONB987. Finally unused C&I loan commitments for banks are obtained from RCONJ547. All the explanatory variables except liquid assets are averages from the first quarter of 2019 through the last quarter of 2019. Total assets and liquid assets of each bank are obtained from the last quarter of 2019.

The reason Call Report data from the first quarter of 2020 is not used is because at the time Covid-19 emergency procedures are first implemented and the data from banks may deviate from the norm. I scale all the explanatory variables and the dependent variable by individual bank's total assets in order to normalize the data. Observations where the PPP loan amount outstanding are many times the total assets of a bank are dropped because there is a possibility of wrong data entry. I use two sets of data. The first contain a total 5086 observations of individual banks existing in the second quarter of 2020 with total assets ranging from $50 million to $3 trillion. I then separate banks based on total assets. Following Li and Strahan (2020), I consider banks with less than $10 billion in total assets as small banks. I obtain 4947 observations. Medium banks have total assets less than $50 billion but higher than $10 billion, I obtain 96 observations. Finally, large banks are those with total assets greater than $50 billion and there are 43 such observations. I use the data containing the small banks and the data containing all banks. Summary statistics of the two sets of data are given in Table 1 and Table 2 respectively. From the table, it is apparent that small banks provided a slightly higher amount of PPP loans/asset and at the same time had slightly higher core deposits and liquid assets and less unused C&I loan commitments when normalized by total assets.

4 Estimation and Results

I use six explanatory variables for the data that have observations from all banks. These are C&I loan commitments for small businesses less than $250,000, C&I loan commitments for small businesses between $250,000 and $1,000,000, C&I loan commitments for large businesses, core deposits, liquid assets and unused C&I loan commitments. For the data that contain information on only small banks, I use all the explanatory variables initially. However, the estimates of the regression function have wild curvatures when plotted which meant that the estimates of the regression function have high bias. As a result for the the latter dataset, I use only five explanatory variables. These are C&I loan commitments for small businesses less than $250,000, C&I loan commitments for small businesses between $250,000 and $1,000,000, core deposits, liquid assets and finally, unused C&I loan commitments. The reason C&I loan commitments for large businesses are dropped is because from the data consisting only of small banks I observe a sufficiently high number of banks with zero values in C&I loan commitments to large businesses.

I use two estimation methods to estimate each dataset. First, I use an Ordinary Least Square (OLS) regression using the "lm" function in R. I test the null hypothesis that this parametric linear model is correctly specified given the data using the parametric regression model specification test by Hsiao et al. (2007) and implemented by the "npcmstest" function within the "np" package written by Li and Racine (2007). The test takes both categorical and continuous data, although I only have continuous data. The naive linear models are rejected by both sets of data. The result of these tests are shown in Table 3.

I then use the "np" package to perform multivariate Kernel regressions on both sets of data. Initially, I obtain the bandwidths for each of the explanatory variables using the "npregbw" function. The Kernel function used defaults to a Gaussian and bandwidths are obtained using least squares cross validation. I use the "npreg" function to obtain the estimates of the fitted model. I use bootstrap to find upper and lower error bounds of the regression estimates at different quantiles of the explanatory variables. I separately graph the marginal effects of each explanatory variable, holding all the other explanatory variables at their median and show the corresponding upper and lower error bounds of the estimates of the regression function in Figure 1 through Figure 6 for observations from all banks and Figure 7 through Figure 11 for observations from only small banks. Finally, I test the significance levels of each of the explanatory variables using the function "npsigtest" in "np" package with 399 bootstrap replications. I use the "Wild" bootstrap method originally proposed by Wu (1986) and Liu (1988). This is because, the observations in both sets of data are not i.i.d and may have heteroskedasticity. The results of the significance tests are shown in Table 4.

Finally, I use semiparametric smooth coefficient kernel regressions in order to compare with the coefficient estimates obtained from the parametric OLS regression. The results of the OLS regressions and the corresponding results from the smooth coefficient Kernel regressions are shown in Table 5 and Table 6 respectively. In the latter case, the coefficients for each of the explanatory variables are calculated using the "npscoefbw" and "npscoef" functions from the "np" package. According to Li and Racine (2007), the "npscoefbw" function computes a bandwidth object for each explanatory variable using least squares cross validation method, with the Kernel function defaulting to a Gaussian. The "npscoef" function then computes a smooth coefficient kernel regression of a one-dimensional dependent variable on *p-variate* explanatory data using the model $Y_i = W_i'\beta(Z_i)+u_i$ where $W_i = (1, X_i')$. Here, $\beta(.)$ is a function that is not specified, however it is assumed that $E(u_i|X_i, Z_i) = 0$. Given that S is the support of Z_i, for any z in S, the function $\beta(Z) = [E(X_iX_i'|Z_i = z)]^{-1}(E(X_iY_i|Z_i = z)$ is estimated by the local constant kernel method. A detailed explanation of the estimation method and the corresponding estimator is provided in Li and Racine (2010).Using the function “npscoef”, I obtain a vector of β values for each explanatory variable, with the vector length corresponding to the number of observations in the data. I also obtain a vector of residuals, with the vector length corresponding to the number of observations in the data. I compute a weighted mean of the β values, with (1 - residuals) as weights. This allows me to obtain comparable β coefficients using a semiparametric approach while the “lm” function in R produces parametric β coefficient estimates.

Comparison between the two sets of β values should be limited to the sign of each of the explanatory variables. This is because, the β coefficient estimates obtained from the linear model are not reliable as the linear parametric model is rejected by the parametric model specification test described earlier. Similarly, the β coefficient estimates from the semiparametric model are a weighted mean and may not reflect the “true β” values. At the same time, it was not possible to use the function “npsigtest” to test the significance levels of the coefficient estimates due to the nature of the semiparametric regression performed by “npscoef”.

It is apparent from Table 4, Table 5 and Table 6 that the nonparametric multivariate Kernel regressions and the semiparametric smooth coefficient Kernel regressions provide higher R^2 values than the linear OLS regressions. Similarly, the residual standard errors are also less for both the nonparametric and semiparametric Kernel regressions. I could not however obtain standard errors for the explanatory variables for either the nonparametric or the semiparametric regressions.

Nevertheless, from the figures, Figure 1 through Figure 11, it is possible to draw inference of the marginal effects of each of the explanatory variables on the amount of PPP loans distributed per bank. Banks which distributed a greater proportion of C&I loans less than $250,000 over the past four quarters are less likely to distribute higher amounts of PPP loans. This may be explained by the fact that such banks are more likely to be located in socioeconomically vulnerable areas and as shown by Wang and Kang (2023), communities with greater economic vulnerability did not receive their proportional share of PPP loans.

However, banks that distributed C&I loans between $250,000 and $1,000,000 over the previous four quarters consistently provide higher amount of PPP loans. This ties in with the findings from Balyuk et al. (2020) and Bartik et al. (2020) who found that banks target loans to their already existing customers and larger borrowers also enjoy early PPP loan access. It is also evident from the figures that the amount of PPP loans disbursed is positively related to core deposits of the banks and unused C&I loan commitments over the past four quarters. These findings are expected and confirmed by existing literature on the relationship between C&I loans disbursed by banks and their core deposits and unused loan commitments. C& I loan commitments to large businesses are not a significant predictor of the amount of PPP loans that the banks disbursed, Li and Strahan (2020) find a similar result, although they use a parametric linear model. This may be because most smaller banks do not distribute large quantities of C& I loans to large businesses as evident from the data. The only surprising finding is that the amount of PPP loans disbursed are negatively related to the liquid assets of the banks. This phenomenon is also evident in the sign of the coefficient estimate for the variable Liquid Assets in both the semiparametric models shown in Table 5 and Table 6. Interestingly, Li and Strahan (2020) also find negative coefficient estimates for Liquid Assets in their parametric model. In other words, banks with higher liquidity disbursed less PPP loan funds, a phenomenon which runs counter to existing literature discussed previously regarding liquid assets and the extent of C&I loan disbursement.

5 Conclusion

Using FDIC Call Report data from June 2020, I measure how much PPP loans/assets were given out by FDIC registered banks. I obtain C&I loan commitments of the banks to large businesses, C&I loans issued for small businesses, core deposits,liquid assets and unused C&I loan commitments over the four quarters of 2019 as predictors for the amount of PPP loans given out by each bank. I use nonparametric multivariate Kernel regressions and semiparametric smooth coefficient Kernel regressions to understand what institutional features lead to a higher amount of PPP loans/asset disbursed per bank and provide significance levels for each of the explanatory variables. C&I loan commitments to large businesses are not a significant predictor of the amount of PPP loans given out by the banks. All other explanatory variables except C&I loan commitments to small businesses of less than $250,000 and liquid assets positively influence the amount of PPP loans that are disbursed by banks. The former phenomenon is discussed in other papers that study how economically vulnerable businesses and communities do not get early access to PPP loans, the same businesses who are likely to receive C&I loans commitments of less than $250,000 in the previous four quarters. The latter phenomenon however is not widely discussed in existing literature and is fertile ground for further research.

References

Allen, K. D. and M. D. Whitledge (2022): "Further evidence on the effectiveness of community banks in the Paycheck Protection Program," *Finance Research Letters*, 47, 102583.

Amiram, D. and D. Rabetti (2020): "The Relevance of Relationship Lending in Times of Crisis," Available at SSRN: https://ssrn.com/abstract=3701587.

Autor, D., L. D. Crane, D. Cho, M. Goldar, B. Lutz, J. K. Montesz, W. B. Peterman, D. D. Ratner, D. V. Vallenas, and A. Yildirmaz (2022): "An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata," *Journal of Economic Perspectives*, 36, 55–80.

Avery, R. B. and A. N. Berger (1991): "Loan commitments and bank risk exposure," *Journal of Banking and Finance*, 15, 173–192.

Avery, R. B. and K. A. Samolyk (2004): "Bank Consolidation and Small Business Lending: The Role of Community Banks," *Journal of Financial Services Research*, 25, 291–325.

Balyuk, T., N. R. Prabhala, and M. Puri (2020): "Indirect Costs of Government Aid and Intermediary Supply Effects: Lessons From the Paycheck Protection Program," NBER Working Paper no.28114.

Barraza, S., M. Rossi, and T. J. Yeager (2020): "The Short-Term Effect of the Paycheck Protection Program on Unemployment," Available at SSRN: https://ssrn.com/abstract=3667431.

Bartik, A. W., Z. B. Cullen, E. L. Glaeser, M. Luca, C. T. Stanton, and A. Sunderam (2020): "The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses," NBER Working Paper no.27623.

Berlin, M. and L. J. Mester (1999): "Deposits and Relationship Lending," *The Review of Financial Studies*, 12, 579–607.

Èlizbar Nadaraya (1964): "On Estimating Regression," *Teoriya Veroyatnostei i Ee Primeneniya*, 9, 157–159.

Fan, J. and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.

Faulkender, M. W., R. Jackman, and S. Miran (2021): "The Job Preservation Effects of Paycheck Protection Program Loans," Available at SSRN: https://ssrn.com/abstract=3767509.

García-Portugués, E. (2023): "Notes for Nonparametric Statistics," https://bookdown.org/egarpor/NP-UC3M/, accessed: 21August, 2023.

Granja, J., C. Makridis, C. Yannelis, and E. Zwick (2020): "Did the Paycheck Protection Program Hit the Target?" NBER Working Paper no.27095.

Harvey, J. and K. Spong (2001): "The Decline in Core Deposits? What can Banks do?" Tech. rep., Federal Reserve Bank of Kansas.

Hayfield, T. and J. S. Racine (2008): "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, 27, 1–32.

Hlavac, M. (2022): *stargazer: Well-Formatted Regression and Summary Statistics Tables*, Social Policy Institute, Bratislava, Slovakia, r package version 5.2.3.

Howell, S. T., T. Kuchler, D. Snitkof, J. Stroebel, and J. Wong (2021): "Racial Disparities in Access to Small Business Credit: Evidence from the Paycheck Protection Program," CESifo Working Paper No. 9345.

Hsiao, C., Q. Li, and J. S. Racine (2007): "A Consistent Model Specification Test with Mixed Categorical and Continuous Data," *Journal of Econometrics*, 140, 802–826.

Kashyap, A. K., R. Rajan, and J. C. Stein (2002): "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking," *The Journal of Finance*, 57, 33–74.

Kim, D. and W. Sohn (2017): "The effect of bank capital on lending: Does liquidity matter?" *Journal of Banking and Finance*, 77, 95–107.

Lester, T. W. and M. D. Wilson (2023): "The Racial and Spatial Impacts of the Paycheck Protection Program," *Economic Development Quarterly*, 4, 1–16.

Li, L. and P. Strahan (2020): "Who Supplies PPP Loans (And Does It Matter)? Banks, Relationships and The Covid Crisis," NBER Working Paper no.28286.

Li, Q. and J. S. Racine (2007): *Nonparametric Econometrics*, Princeton: Princeton University Press.

——— (2010): "Smooth Varying-Coefficent Estimation and Inference for Quantitative and Qualitative Data," *Econometric Theory*, 26, 1607–1637.

Liu, R. Y. (1988): "Bootstrap Procedures under some Non-I.I.D. Models," *The Annals of Statistics*, 16, 1696–1708.

Nguyen, H.-L. Q. (2019): "Are Credit Markets Still Local? Evidence from Bank Branch Closings?" *American Economic Journal: Applied Economics*, 11, 1–32.

Nguyen, N. T. H. and J. R. Barth (2020): "Community Banks vs. Non-Community Banks: Where is the Advantage in Local Small Business Funding?" *Atlantic Economic Journal*, 48, 161–174.

Wand, M. P. and M. C. Jones (1995): *Kernel Smoothing*, London: Chapman & Hall.

Wang, Q. and W. Kang (2023): “Small businesses and government assistance during COVID-19: Evidence from the paycheck protection program in the U.S.” *Environment and Planning A: Economy and Space*, 0, 0.

Wu, C. (1986): “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis,” *The Annals of Statistics*, 14, 1261–1295.

Table 1: Summary statistics of the Dependent variable and the Explanatory variables from all banks. All the variables are normalized by total assets of each bank. This table is made using the "stargazer" package by Hlavac (2022)in R.

Statistic	N	Mean	St. Dev.	Min	Max
Amount of PPP loans	5,086	0.057	0.074	0.000	0.849
C&I Less than $250,000	5,086	0.016	0.023	0.000	0.792
C&I Less than $1,000,000	5,086	0.014	0.015	0.000	0.244
C&I More than $1,000,000	5,086	0.020	0.053	0.000	0.831
Core Deposits	5,086	0.258	0.147	0.000	0.895
Liquid Assets	5,086	0.283	0.167	0.000	0.999
Unused C&I Loan Commitments	5,086	0.029	0.032	0.000	0.264

Table 2: Summary statistics of the Dependent variable and the Explanatory variables from small banks with assets less than $10 billion. All the variables are normalized by total assets of each bank. This table is made using the “stargazer” package by Hlavac (2022)in R.

Statistic	N	Mean	St. Dev.	Min	Max
Amount of PPP loans	4,947	0.058	0.074	0.000	0.849
C&I Less than $250,000	4,947	0.016	0.023	0.000	0.792
C&I Less than $1,000,000	4,947	0.014	0.015	0.000	0.244
Core Deposits	4,947	0.263	0.146	0.000	0.895
Liquid Assets	4,947	0.284	0.167	0.000	0.999
Unused C&I Loan Commitments	4,947	0.028	0.032	0.000	0.264

Table 3: This table shows how the parametric null OLS model specification is rejected for both the sample of all banks and the sample of small banks using the test demonstrated in Hsiao et al. (2007).

	Dependent variable:	
	Amount of PPP loans	
	(1)	(2)
Observations	5,086	4.947
IID Bootstrap Replications	399	399
Number of regressors	6	5
Test Statistic 'Jn'	7.258021***	8.625261***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 4: This table shows $p-values$ from the nonparametric multivariate Kernel regressions and the corresponding significance of the explanatory variables obtained using the function "npsigtest" in the "np" package. The first columns shows results from observations of all banks while the second column shows observations from banks with total assets less than $10 billion.

	Dependent variable:	
	Amount of PPP loans	
	(1)	(2)
Less than $250,000 in C&I Loans	0.005	0.025
Less than $1,000,000 in C&I Loans	< 0.0001	< 0.0001
More than $1,000,000 in C&I Loans	0.108	–
Core Deposits	< 0.0001	0.013
Liquid Assets	< 0.0001	< 0.0001
Unused C&I Loan Commitments	< 0.0001	< 0.0001
Observations	5,086	4,947
Kernel Regression Estimator	Local-Linear	Local-Linear
Bandwidth Type	Fixed	Fixed
R^2	0.295	0.225
Residual Std. Error	0.062	0.065
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 5: This table compares the parametric OLS regression in column 1 with the semi-parametric smooth coefficient Kernel regression in column 2 based on observations from all banks.

	Dependent variable:	
	Amount of PPP loans	
	(1)	(2)
Less than \$250,000 in C&I Loans	-0.083^{*}	-0.556
	(0.047)	–
Less than \$1,000,000 in C&I Loans	0.967^{***}	1.796
	(0.077)	–
More than \$1,000,000 in C&I Loans	-0.117^{***}	0.0341
	(0.021)	–
Core Deposits	0.013^{*}	0.025
	(0.007)	–
Liquid Assets	-0.037^{***}	-0.022
	(0.006)	–
Unused C&I Loan Commitments	0.720^{***}	0.703
	(0.034)	–
Constant	0.033^{***}	0.026
	(0.003)	–
Observations	5,086	5,086
R^2	0.189	0.477
Adjusted R^2	0.188	-
Residual Std. Error (df = 5079)	0.066	0.054
F Statistic (df = 6; 5079)	197.166^{***}	-
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6: This table compares the parametric OLS regression in column 1 with the semiparametric smooth coefficient Kernel regression in column 2 based on observations from small banks with assets less than \$10 billion.

	Dependent variable:	
	Amount of PPP loans	
	(1)	(2)
Less than \$250,000 in C&I Loans	-0.112^{**}	-0.749
	(0.048)	–
Less than \$1,000,000 in C&I Loans	0.864^{***}	2.421
	(0.078)	–
Core Deposits	0.021^{***}	0.044
	(0.007)	–
Liquid Assets	-0.038^{***}	-0.037
	(0.006)	–
Unused C&I Loan Commitments	0.696^{***}	0.506
	(0.034)	–
Constant	0.032^{***}	0.026
	(0.003)	–
Observations	4,947	4,947
R^2	0.185	0.595
Adjusted R^2	0.184	-
Residual Std. Error (df = 4941)	0.067	0.048
F Statistic (df = 5; 4941)	224.489^{***}	-
Note:	*p<0.1; **p<0.05; ***p<0.01	

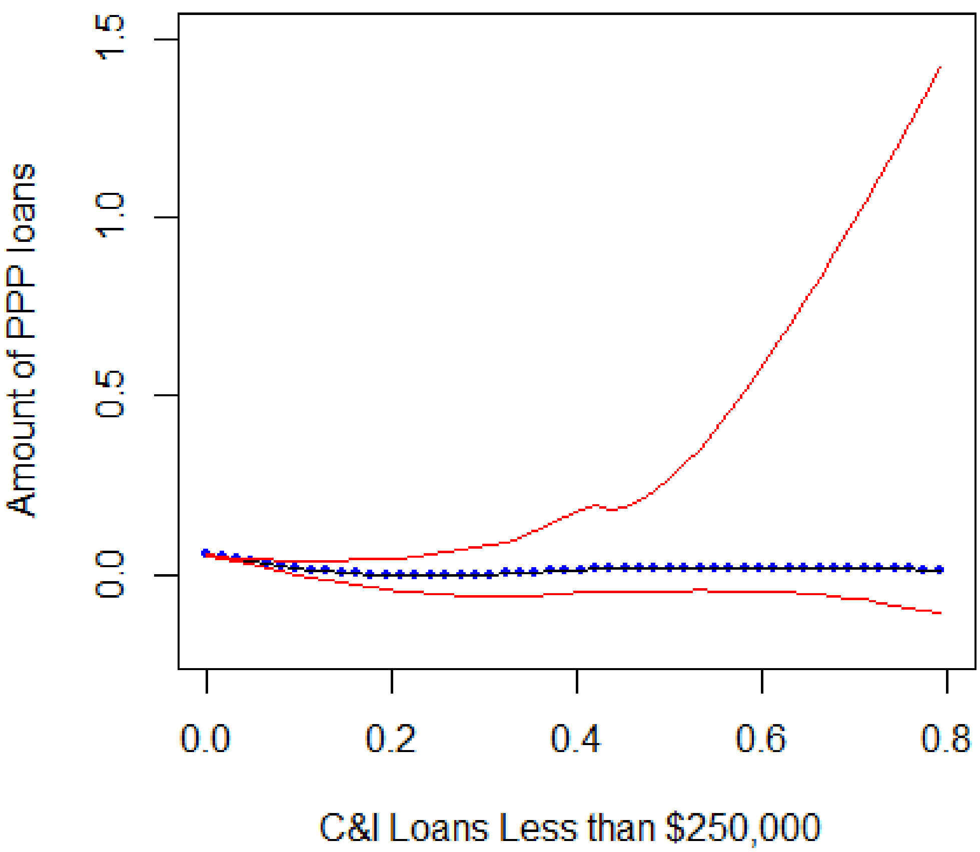

Figure 1: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "C&I loans less than $250,000". The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

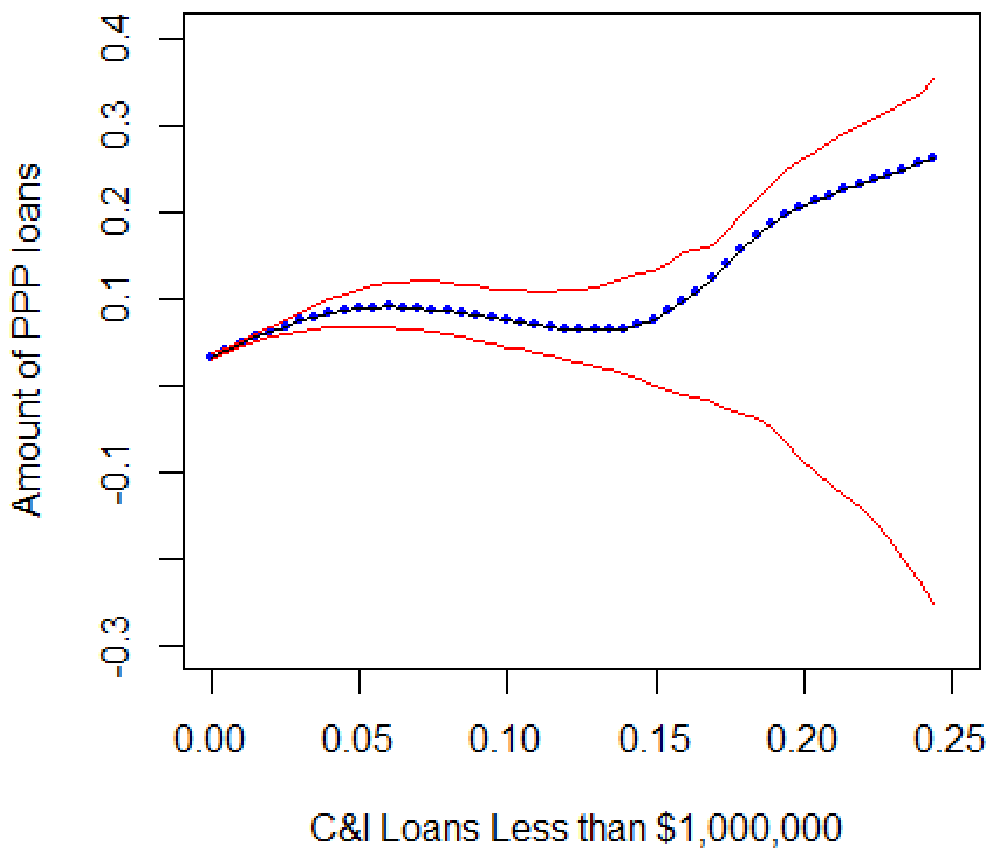

Figure 2: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "C&I Loans less than $1,000,000". The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

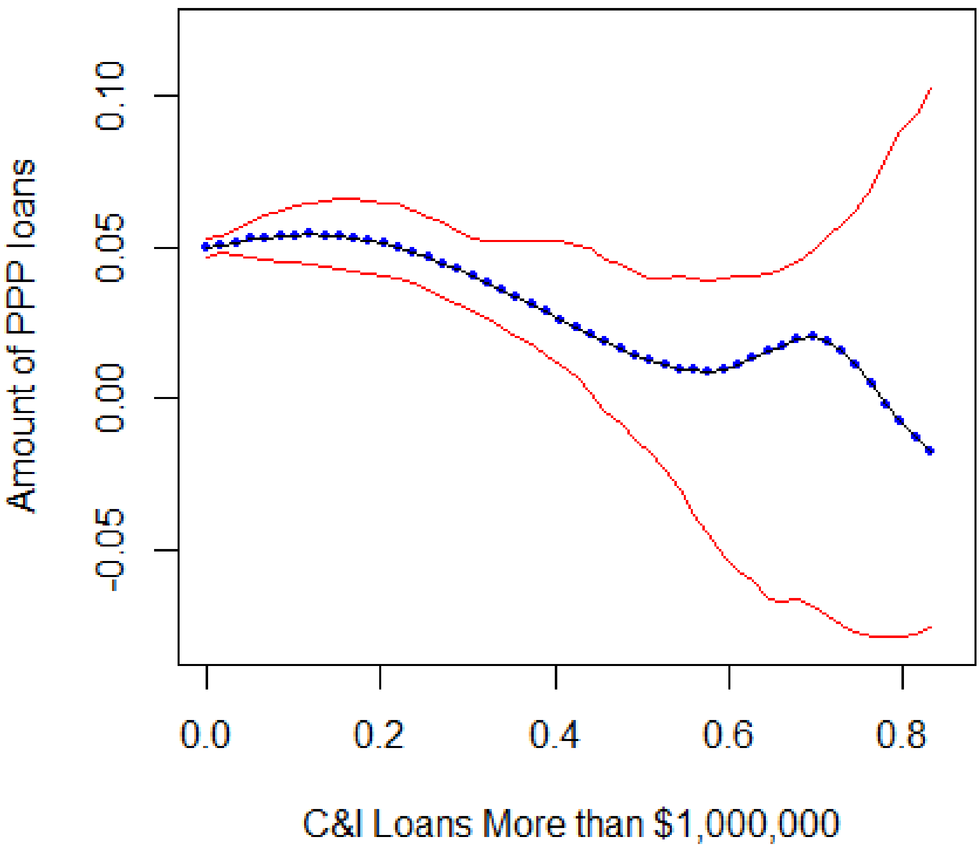

Figure 3: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "C&I Loans more than $1,000,000". The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

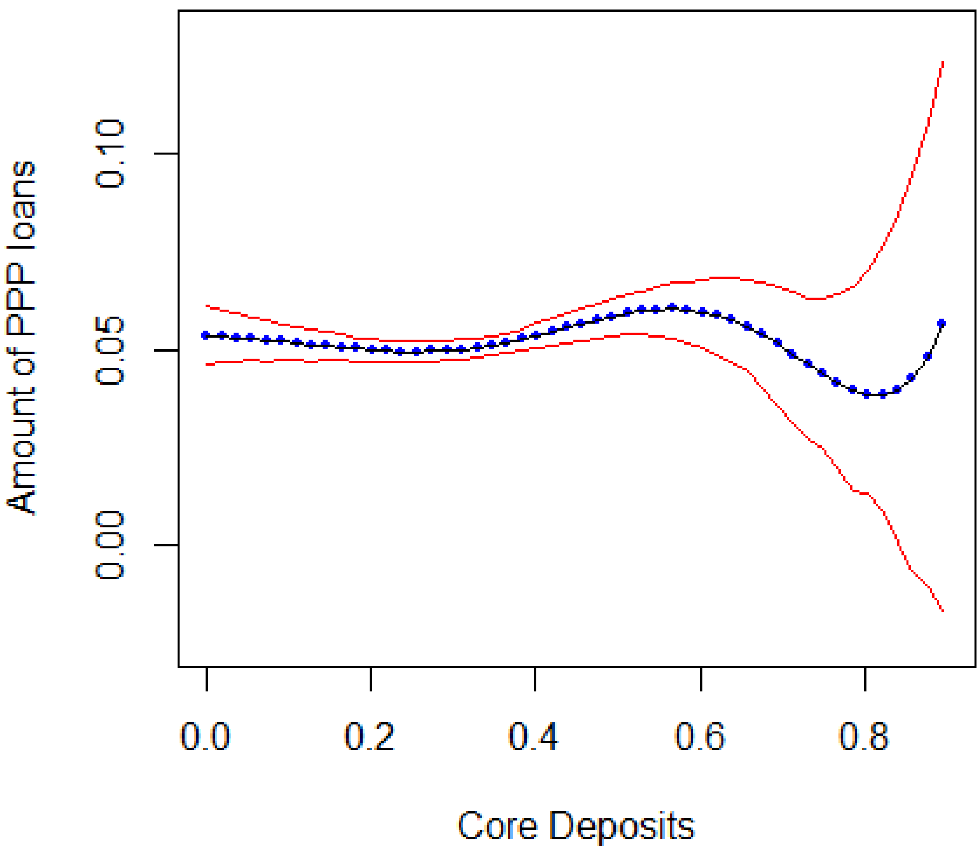

Figure 4: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, “Core Deposits”. The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

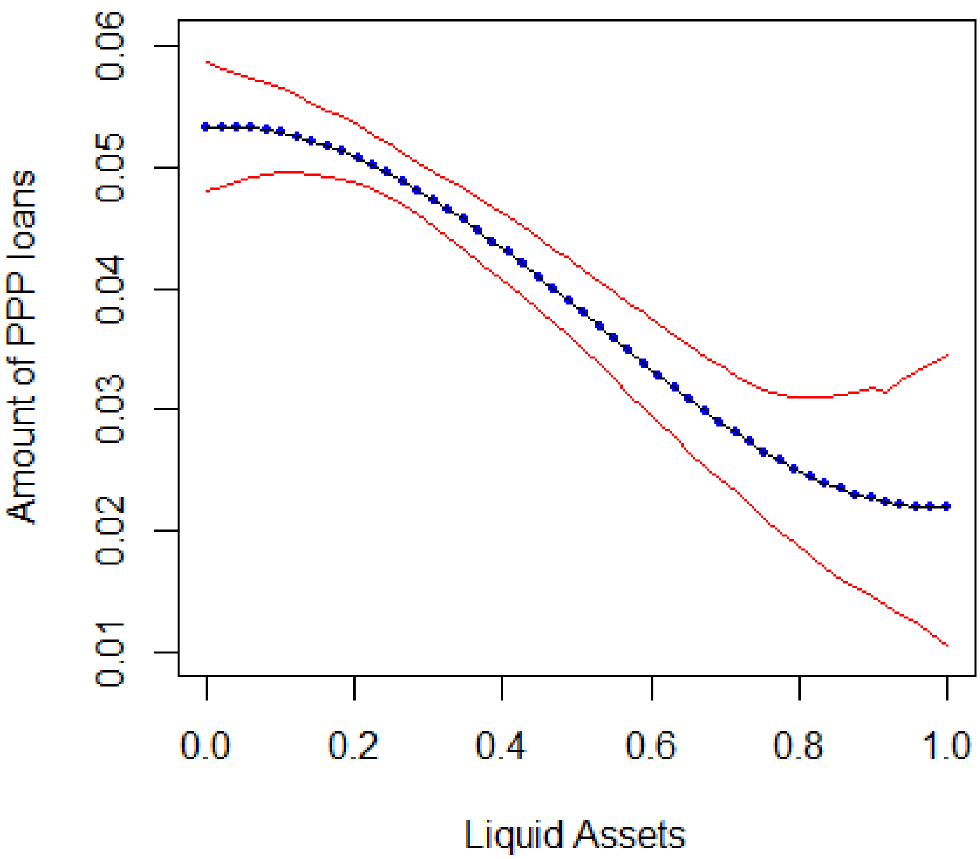

Figure 5: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, “Liquid Assets”. The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

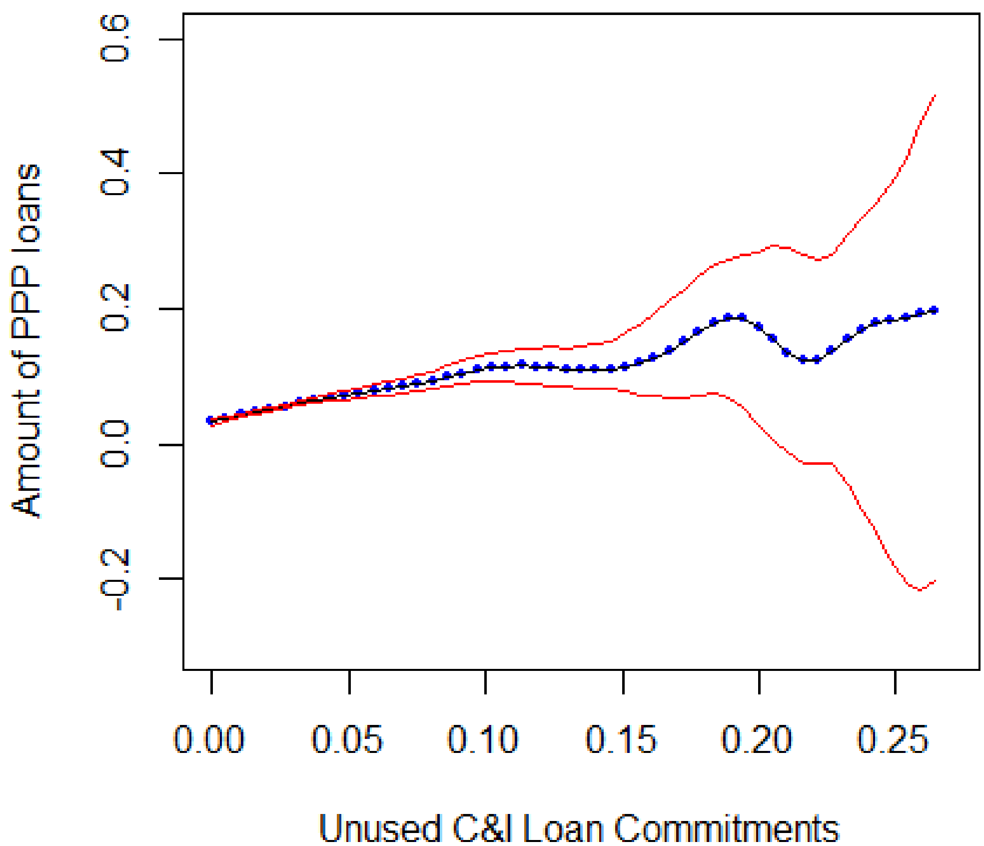

Figure 6: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, “Unused C&I Loan Commitments”. The error bounds are calculated using a bootstrap procedure. Data are from observations are from all banks.

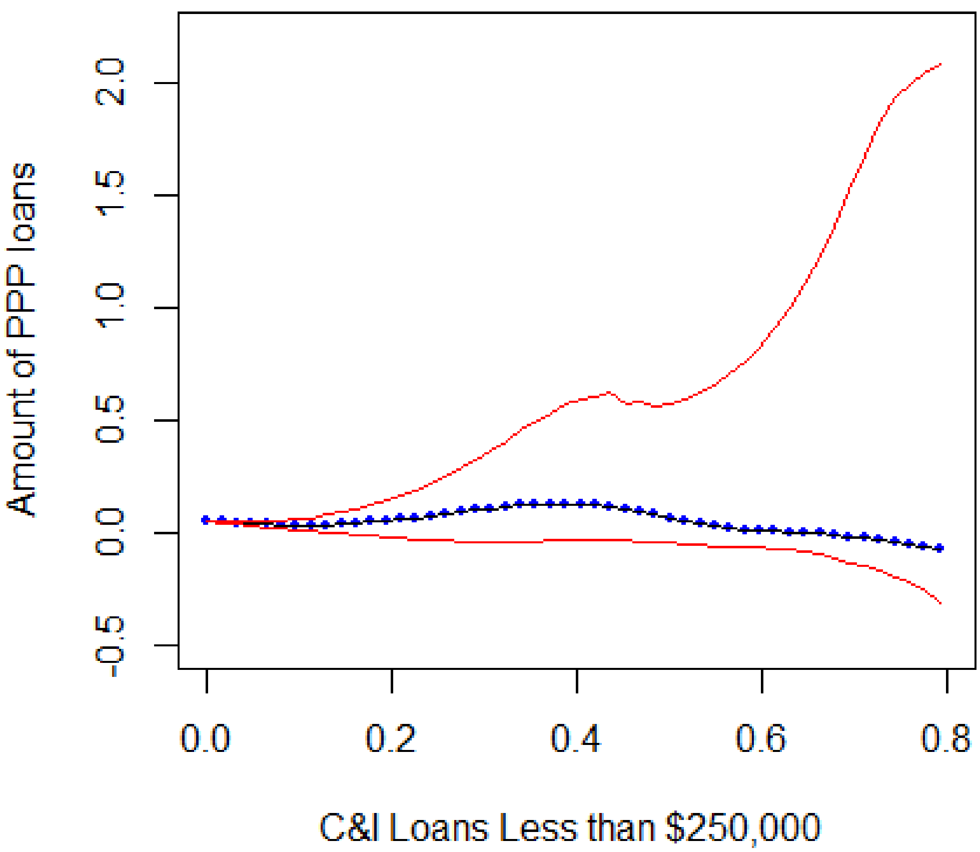

Figure 7: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "C&I loans less than $250,000". The error bounds are calculated using a bootstrap procedure. Data are from observations are from small banks with total assets less than $10 billion.

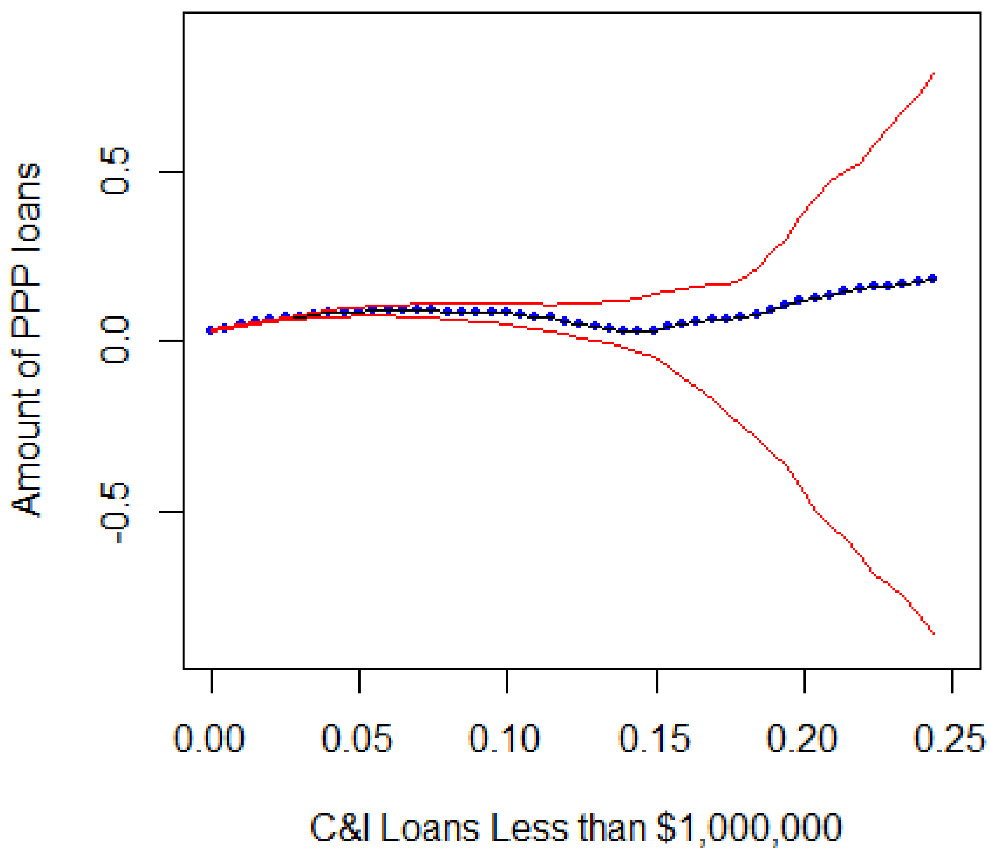

Figure 8: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, “C&I loans less than $1,000,000”. The error bounds are calculated using a bootstrap procedure. Data are from observations are from small banks with total assets less than $10 billion.

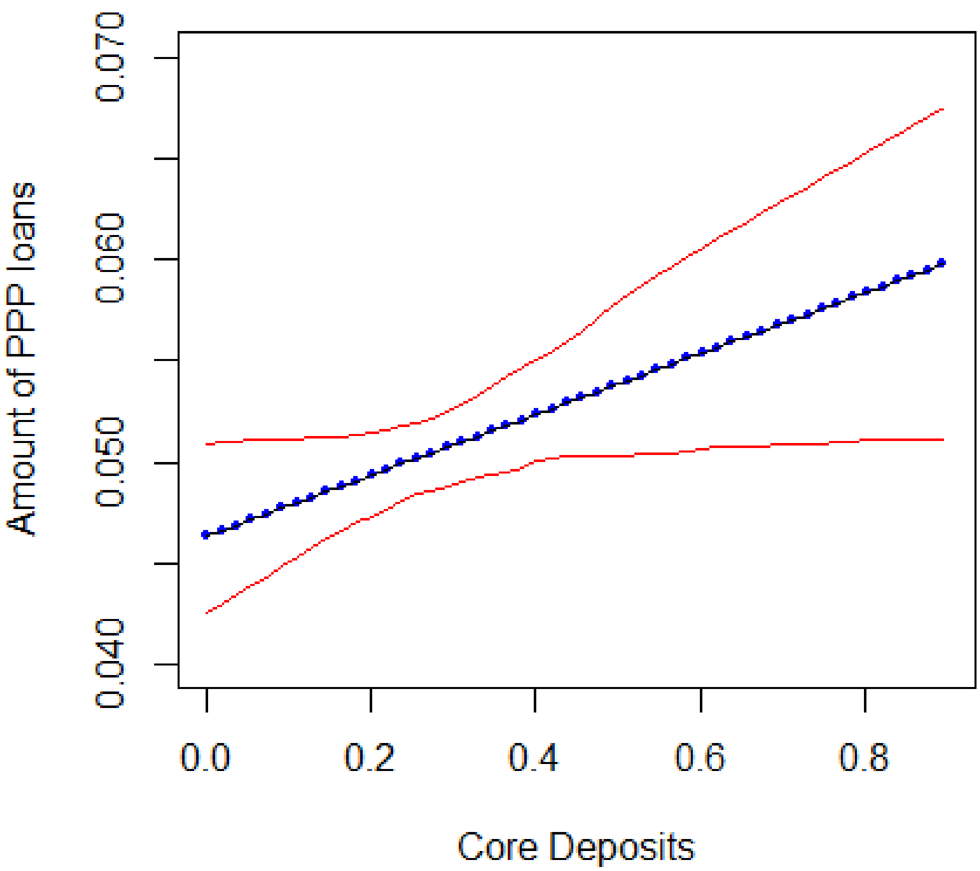

Figure 9: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "Core Deposits". The error bounds are calculated using a bootstrap procedure. Data are from observations are from small banks with total assets less than $10 billion.

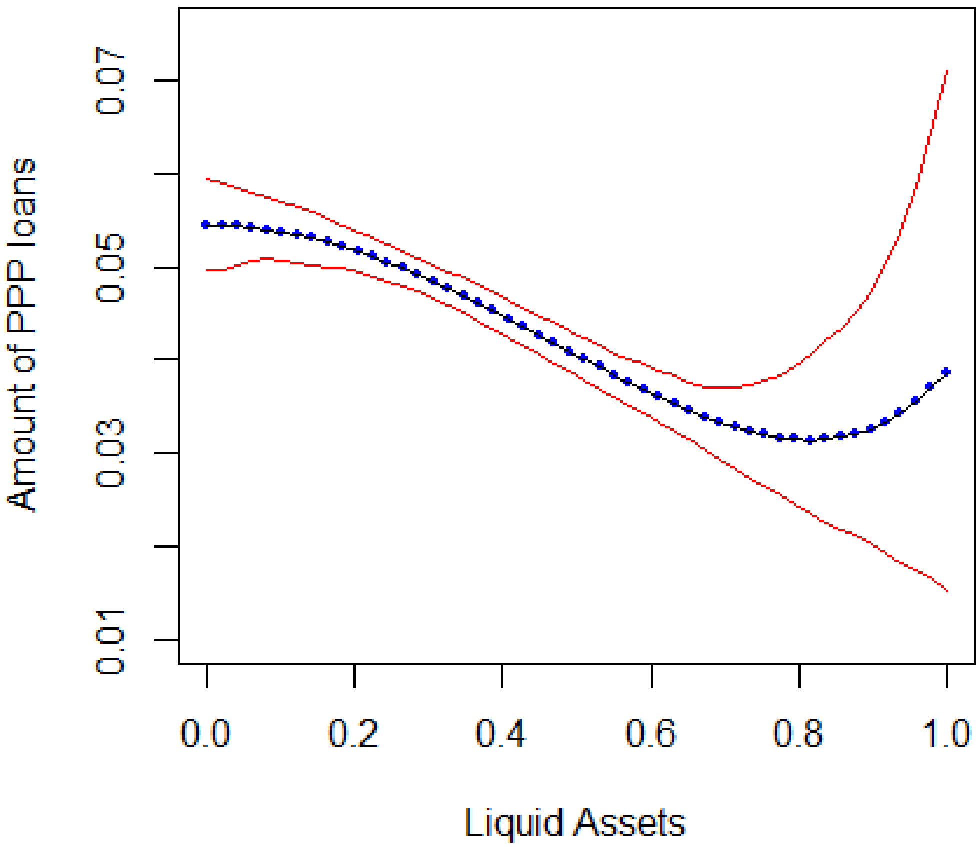

Figure 10: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, "Liquid Assets". The error bounds are calculated using a bootstrap procedure. Data are from observations are from small banks with total assets less than $10 billion.

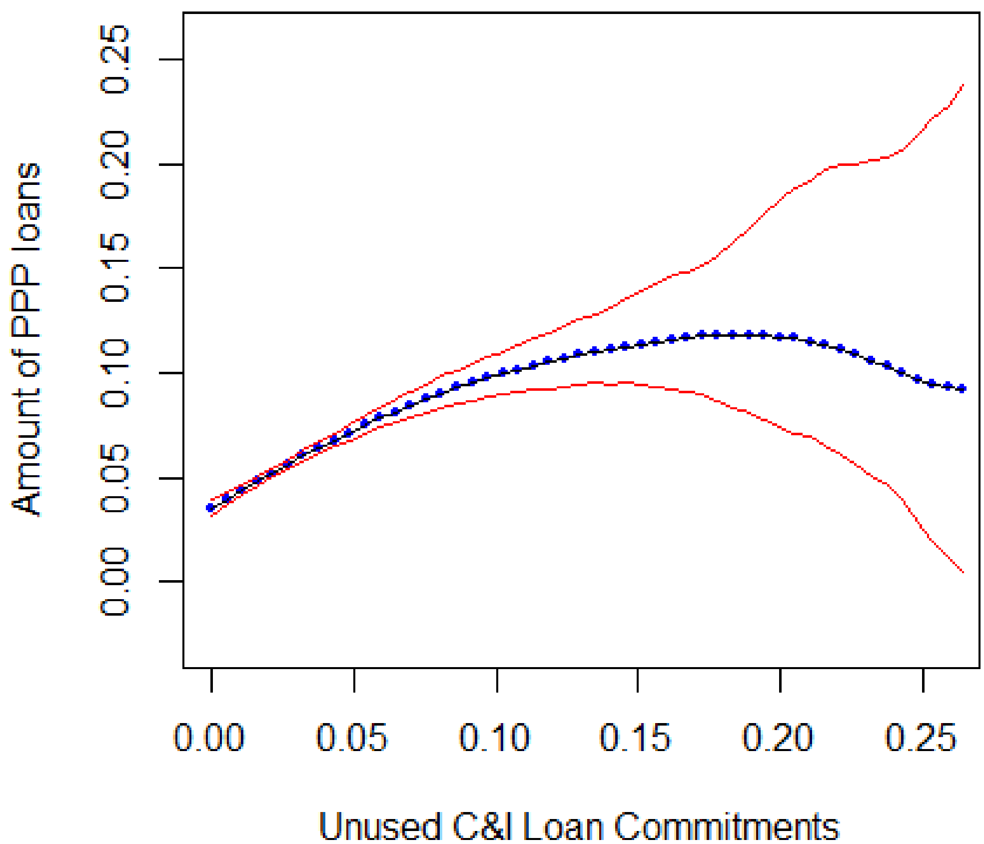

Figure 11: Upper and lower error bounds of the fitted estimates of the regression function at different quantiles of the dependent variable, “Unused C&I Loan Commitments”. The error bounds are calculated using a bootstrap procedure. Data are from observations are from small banks with total assets less than $10 billion.