



# Hispanic brokers and borrowers: The effect of language affinity on the price of home mortgages<sup>☆</sup>



Chau Do<sup>a,\*</sup>, Arturo Gonzalez<sup>b</sup>

<sup>a</sup> Office of the Comptroller of the Currency, 400 7th St SW, Washington, DC 20219, United States

<sup>b</sup> Board of Governors of the Federal Reserve System, 20th St & Constitution Ave NW, Washington, DC 20551, United States

## ARTICLE INFO

### Article history:

Received 15 February 2014

Received in revised form 30 September 2014

Accepted 5 November 2014

Available online 13 November 2014

### JEL classification:

J10

J7

R18

D4

### Keywords:

Hispanic

Language

Mortgage

## ABSTRACT

Our study focuses on whether mortgage prices for Hispanic borrowers in areas of limited English fluency depend on the broker's ethnicity. While we find that Hispanic borrowers in areas where the majority of Hispanics are not fluent in English pay higher prices overall, mortgage prices are lower if the loans are originated by Hispanic brokers relative to non-Hispanic white brokers. This effect is found only in fixed-rate mortgage loans and for low/no-documentation loans. Nevertheless, our results cannot be easily explained by different levels of market competition or borrower characteristics. Our results are consistent with other empirical studies that find that language barriers carry additional costs. We note that our conclusions may be affected by the accuracy of our proxy measurements of English and Spanish fluency.

Published by Elsevier B.V.

## 1. Introduction

The wide disparities in the mortgage market among racial and ethnic groups continue to persist despite government programs and regulations designed to promote minority homeownership. Discrimination, imperfect markets, and financial literacy have been among the most widely cited explanations. There has been limited research, however on the effects of cultural affinity in mortgage lending. In contrast to other discrimination theories, the theory of cultural affinity emphasizes the role of a lender's race/ethnicity in explaining racial disparities. Reflecting the two distinct versions of discrimination (i.e., taste-based and statistical), the literature on cultural affinity has developed two distinct versions of the theory. The *taste-based* form of cultural affinity expands on Becker's (1994) seminal work and assumes that lenders have preferences for members of their own race/ethnicity. Lenders will favor marginal applicants with the same race/ethnicity and penalize marginal applicants that do not have the same race/ethnicity as they do. The two significant implications from the theory are: 1) approved applicants of a different race/ethnicity than that of the lender will be more creditworthy than that of approved

applicants of the same race/ethnicity and 2) borrowers will select lenders of the same race/ethnicity.

The *common bond* formulation of cultural affinity proposed by Calomiris et al. (1994) expands on the theory of statistical discrimination proposed by Arrow (1971) to argue that lenders find it easier to evaluate the creditworthiness of applicants with whom they have a common background or cultural affinity. As such, lenders that share a common cultural background with the applicant find that the costs of gathering additional information about the true creditworthiness of potential borrowers are lower. There are two important implications of the common bond form of cultural affinity. First, marginally uncreditworthy applicants may choose to seek lenders of a different background to capitalize on the errors<sup>1</sup> that are more likely to occur from lenders of a different background, while marginally creditworthy applicants may choose to seek lenders of the same background to signal their creditworthiness (Bostic, 2003). Secondly, the effect of cultural affinity may vary according to the value of the signal (Longhofer, 1996). For example, if the signal is imprecise relative to the variation of the applicant pool, the effect of cultural affinity may not be significant. However, if the signal is more precise relative to the variation of the applicant pool, the effect of cultural affinity may be significant.

<sup>☆</sup> The views in this paper are those of the authors and do not reflect those of the Office of the Comptroller of the Currency, the Department of Treasury, or the Federal Reserve Board.

\* Corresponding author.

E-mail addresses: [chau.do@occ.treas.gov](mailto:chau.do@occ.treas.gov) (C. Do), [arturo.gonzalez@frb.gov](mailto:arturo.gonzalez@frb.gov) (A. Gonzalez).

<sup>1</sup> Errors may stem from the valuation of their credit quality.

Evidence of the importance of a lender's race or ethnicity can be found in related literature (Hunter and Walker, 1996; Bostic, 2003). However, the extant literature focuses only on the effect of the similarity of the borrower and lender's race/ethnicity on the loan's disposition; it does not assess the impact of cultural affinity on other aspects of the loan lifecycle, such as the loan's price or loan performance, or assess whether certain aspects of cultural affinity, such as a common language, matter.

However, there is evidence which suggests that a common language is an important aspect in market transactions (Lang, 1986). Research has shown that Hispanics prefer to read and speak in Spanish, Hispanics with limited English proficiency prefer using a Hispanic broker (Toussaint-Comeau, 2000), and insufficient English skills are an obstacle in gaining access in the housing market (Ratner, 1996; Toussaint-Comeau, 2000; Borjas, 2002; Mosaica, 2004; Cortes et al., 2006; Haurin and Rosenthal, 2009; Kochhar et al., 2009).

Consistent with these findings for Hispanics, banks have made a concerted effort in the last several years to attract more business from Hispanic borrowers in particular by marketing in Spanish and advertising in Hispanic media (Diaz McConnel and Marcelli, 2007; Jones, 2012). While research suggests that Hispanics who identify themselves as primarily reading and speaking Spanish prefer having documents available in Spanish, the translation of documents in Spanish, as well as the supply of Spanish-speaking loan officers working in branches, has been limited, presumably because of cost and supply constraints.<sup>2</sup> Borrowers can circumvent supply constraints by bringing individuals to explain the process in their native language, but this necessarily entails increased costs (directly or indirectly), and it does not fully address the issue of how effectively the borrower can negotiate a lower mortgage price if language barriers exist.<sup>3</sup>

This paper uses data from a major subprime lender and investigates the effect of language on mortgage prices. Specifically, we assess whether mortgage pricing for Hispanic borrowers that live in areas where a majority of Hispanics are limited English proficient (LEP) depends on the broker's ethnicity. We find that Hispanic borrowers that live in areas where Hispanics are primarily LEP have a lower annual percentage rate (APR) when the loan is originated by a Hispanic broker than when the loan is originated by a non-Hispanic white broker. Because our results are driven by fixed rate loans, we cannot rule out unobservable characteristics that are correlated with Hispanic borrowers that live in areas where Hispanics are primarily LEP. However, we do not find differences in the probability of default across Hispanic borrowers in areas where Hispanics are primarily LEP and in areas where Hispanics are not primarily LEP. Additionally, our robustness checks suggest that the lower price is neither reflective of lower credit risk of Hispanic borrowers in areas where Hispanics are predominantly LEP nor of the greater supply of Hispanic brokers. For example, we do not find lower APRs for other minorities or non-Hispanic white borrowers that live in areas where Hispanics are predominantly LEP. This suggests that the effects found may not be driven by unobservable characteristics that are correlated with areas where Hispanics are predominantly LEP.

Our study contributes to the existing literature in several ways. First, we expand the mortgage pricing literature to encompass language affinity as an important and plausible theory affecting pricing disparities among racial and ethnic groups. While loan disposition is an important outcome in assessing constraints in credit supply, there are other important aspects, such as pricing and performance, which provide a fuller picture of the homeownership gap across races and ethnic groups. Additionally, the results shed light on the common bond form of cultural affinity in mortgage lending. Our results show that Hispanic brokers charge higher

<sup>2</sup> Different states have different laws regulating whether Spanish documents can be substitutes for English mortgage documents, whether all or partial documents can be translated, and whether there are requirements that lenders offer documents in the customer's native language.

<sup>3</sup> Further, if one were to assume that a borrower accompanied by an English speaker is a perfect substitute for a borrower that speaks English fluently, our estimates will be biased towards zero in trying to assess the effect of sharing a common language on mortgage prices.

prices to Hispanic borrowers overall but lower prices to Hispanic borrowers that live in areas where Hispanics are predominantly LEP. Additionally, the lower prices are only present for particular types of loans (i.e., home purchase and no/limited documentation loan products) where the value of language or the signal of a borrower's creditworthiness may be more important. Presumably, if lender discrimination is solely based on taste-based preferences, effects on the commonalities of race and language should reinforce each other and are independent of selection effects into certain loan products.<sup>4</sup> Lastly, from a policy perspective, given the growth of non-English speakers due to immigration, the paper suggests that there may be a value to multilingual education or provision of services in multiple languages.<sup>5</sup>

There are a number of potential shortcomings that may influence our results and conclusions. Data limitations necessitate the reliance of some assumptions of broker's and borrower's language fluency. Secondly, while we explore other explanations that drive these results which are consistent with language affinity, we cannot completely rule out other interpretations. In particular, we find evidence that our results are driven by fixed rate loans for which there is no obvious reason why the effect may not be found in ARM loans as well. As such, if our proxies and assumptions are not valid, the strength of our conclusions is weakened. The rest of the paper is organized as follows. Section 2 describes the empirical methodology and data. Section 3 discusses the results and robustness checks; Section 4 concludes the study.

## 2. Empirical strategy

To first motivate our empirical strategy, it is easiest to consider a simple mortgage market model, where the equilibrium price,  $p_m$ , is the price at which a borrower finds a lender willing to supply the mortgage with the characteristics demanded.

$$p_m = f(D, S). \quad (1)$$

The borrower's demand,  $D$ , can be defined as a function of several factors,  $D = D(\text{credit constraints, borrower preferences, borrower characteristics})$ , while the lender's cost function,  $S$ , can be defined by market factors as well as inputs into the provision of a mortgage,  $S = S(\text{market competition, lender preferences, costs})$ . The value of speaking the borrower's native language may enter in the demand function directly as part of borrower preferences. Because language may influence the ability to not only negotiate and understand the financial terms and conditions but also the desire to signal the borrower's creditworthiness, it may enter indirectly through borrower characteristics as well. Language may also enter directly in the supply function as part of the direct costs of offering language services (i.e., extra cost for producing translated documents in a foreign language, hiring multilingual loan officers, etc.) or lender preferences, as well as indirectly in the lender's costs of the time and effort needed to infer the borrower's creditworthiness.<sup>6</sup>

Eq. (1) highlights that it is difficult, a priori, to predict the impact of language as it directly and indirectly affects the demand and supply function. At the most basic level, borrowers and lenders may simply have a preference for those that speak the same language. This suggests that borrowers are willing to pay more for lenders that speak the same language while lenders may provide discounts to borrowers because of taste-based preferences. Which factors prevail determines whether the equilibrium price is higher or lower for borrowers and lenders that share the same language, *ceteris paribus*.

<sup>4</sup> We note that one can argue that selection effects found in certain loan products may be reflective of variation in taste-based preferences by the borrower that may be correlated with creditworthiness. However, lender's preferences should not necessarily vary according to loan products. As such, the equilibrium price could vary according to differences in demand.

<sup>5</sup> Assuming the government's objective is to increase minority homeownership.

<sup>6</sup> Similarly, a common race/ethnicity may enter both the demand and supply function either because of taste based preferences or because of the value of the signal that it can provide.

Alternatively, if a common language can signal better the borrower's creditworthiness, marginally qualified borrowers may select lenders of the same race/ethnicity while marginally unqualified borrowers may select lenders of a different race/ethnicity. The effect of sorting complicates the equilibrium price and the ability to disentangle the true effect of the costs to language barriers.

For expository ease, we ignore the effects of sorting for now and assess how language and race may impact mortgage prices by our empirical regression:

$$APR_{ist} = \beta_1 + \beta_2 H_i^L + \beta_3 LEP_i^B + \beta_4 (LEP_i^B \times H_i^L) + X_i' \delta + S_s' \vartheta + T_t' \gamma + \varepsilon_i \quad (2)$$

where  $APR_{ist}$  is the price of loan  $i$  that Hispanic borrowers<sup>7</sup> receive in year  $t$  in state  $s$ .  $H_i^L$  is an indicator variable equal to 1 if the broker that originated the loan is Hispanic and 0 if the broker that originated the loan is non-Hispanic white.  $LEP_i^B$  is an indicator variable, equal to 1 if the borrower of the loan lives in an area where Hispanics are primarily limited English proficient.  $X_i$  is a vector of loan and borrower characteristics that may affect the APR (i.e., marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, condominium indicator,<sup>8</sup> multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator).  $S_s \vartheta$  and  $T_t \gamma$  are fixed effects for state and time, respectively. We are primarily interested in the estimated coefficient of  $\beta_4$ , which indicates whether or not sharing the same language affects the APR.

## 2.1. Data

To accurately measure the effect of language barriers or conversely, language affinity, data would ideally include information on the borrower's and broker's fluency in English and Spanish and a random assignment of brokers with different levels of English and Spanish fluency to borrowers with different levels of English and Spanish fluency. Unfortunately, we do not have such a natural experiment. However, we have detailed information on broker and borrower characteristics that we use to proxy for Spanish and English fluency. We describe our data and data construction below.

Data on the loan, borrower, and broker characteristics are based on loans originated by New Century from 2004–2006. These data contain information on APR, the borrower's race/ethnicity, the location of the property, and the broker that originated the loan. The comparability of this data with that of HMDA data is discussed in Appendix 1. We define a borrower's race and ethnicity based on the primary applicant's race and ethnicity, regardless of the presence of a co-applicant.<sup>9</sup> Because we only have information on the broker's first and last names, we merge the mortgage data with the US Census Bureau's 2000 surname list that contains surnames and the percentage of instances that the surnames are used by a certain race or ethnicity. From these merged data, we are able to infer the broker's ethnicity by their surname.<sup>10</sup> This approach is commonly done in other studies when surname is used to obtain an individual's racial/ethnic information (Elliot et al., 2009).

<sup>7</sup> For robustness, we replicate Eq. (2) for a sample of non-Hispanic Asian, non-Hispanic white borrowers, and non-Hispanic black borrowers to assess whether there are similar effects. See columns (2)–(4) of Table 2.

<sup>8</sup> Indicator variable for whether the property is a condominium or not.

<sup>9</sup> In Table 5, we also include robustness checks that exclude joint applications that are not Hispanic.

<sup>10</sup> The US Census Bureau released a national tabulation of surnames classified by self-reported race/ethnicity based on almost 270 million individuals with valid surnames. Each surname with 100 or more occurrences nationally is provided publicly, along with the percentage of times that it is linked with an individual of a particular race or ethnicity. We identify Hispanic surnames if the percentage of times that the surname is from a Hispanic individual is at least 80%. See <http://www.census.gov/genealogy/www/freqnames2k.html> for more information.

We also proxy for the borrower's English fluency by incorporating language information at the ZIP code level from the 2006–2010 American Community Survey (ACS) 5-year Summary File. The data contain area-level counts of the language(s) spoken and English language ability of persons 5 years and older. From these data, we calculate the percentage of Hispanics within a ZIP code that speak a foreign language and do not speak English very well.<sup>11,12</sup> We classify Hispanic borrowers that live in an area where Hispanics are primarily LEP if the percentage of Hispanics in the area that do not speak English very well is greater than or equal to 50%. This is commonly done in research when there is no individual-level data available for English fluency (Borjas, 1992, 2002; Bertrand et al., 2000; Toussaint-Comeau, 2000; Berndt et al., 2012). Studies that use this approach find a positive correlation between the externalities attributable to ethnic enclaves and an outcome, such as home ownership, earnings, or education (Borjas, 1992, 2002; Bertrand et al., 2000; Toussaint-Comeau, 2000; Berndt et al., 2012).

However, using geographic concentration of Spanish speakers where the purchased house is located is a possible source of measurement error. This measurement error can lead to the mis-classification of the variable of interest, and therefore bias the estimates (Lewbel, 2007). Although we consider alternative cut-offs to define LEP borrowers, any potential inherent measurement error in this variable is still introduced to all individuals.<sup>13</sup>

We limit our sample from which to estimate our model to Hispanic borrowers who purchased a 30-year, first-lien loan to ensure that the population is relatively homogenous and that the loan prices amortized to the full term reflect the credit risk and transaction costs associated with processing and originating the loan. Missing observations on APR and where we cannot adequately ensure that the broker's surname is Hispanic or non-Hispanic white are dropped.<sup>14,15</sup> These conditions yield a total of 21,056 Hispanic borrowers.

## 2.2. Summary statistics

Table 1 presents summary statistics for the Hispanic sample in our data with non-missing observations.<sup>16</sup> About 35% of Hispanics live in areas where Hispanics are primarily limited-English proficient. There is some evidence of self-selection into certain types of brokers; about 75% of Hispanic borrowers that live in areas where Hispanics are primarily LEP use Hispanic brokers while 67% of Hispanic borrowers that live in areas where Hispanics are primarily EP use Hispanic brokers. Interestingly, Hispanic borrowers that live in areas where Hispanics are primarily LEP receive an APR that is 2.8 basis points higher despite having higher FICO scores. The difference in means (or proportions) between the two groups of borrowers for a majority of the characteristics are statistically significant except for whether the residence is a condominium or not, years on the job, and years in current residence. However, even though the mean characteristics for the two groups of borrowers are statistically

<sup>11</sup> Since the ACS Summary File is organized at the Census-tract level, while the mortgage data is organized at the ZIP code, the data are merged using the crosswalk provided by Master Area Block Level Equivalency (MABLE) file at the Missouri Census Data Center. See <http://mcdc.missouri.edu/webrepts/geography/ZIP.resources.html> (accessed Oct 9, 2012).

<sup>12</sup> Proficiency is typically defined as speaking English well, very well or only English (Chiswick and Miller, 1995; Zavadny, 2000; Dávila and Mora, 2001; Gonzalez, 2013). However, our definition differs slightly from the standard definition of English proficiency because our data only contain counts of persons that speak English very well or English only, in contrast to the five categories that are found in the micro level ACS data. As such, our measure may be a broader measure of defining limited English proficient areas.

<sup>13</sup> Results based on various cut-offs are similar and presented in Appendix Table 2.

<sup>14</sup> Using other cut-offs, results are similar.

<sup>15</sup> As a robustness check, we exclude female brokers in our analysis to ensure that the surname more accurately reflects the individual's ethnicity rather than the spouse's ethnicity.

<sup>16</sup> Appendix 1 compares our sample of Hispanic borrowers to Hispanics in the HMDA 2004–2006 data that are merged with the OCC Mortgage Metrics (MM) to obtain borrower's creditworthiness characteristics. The differences in our sample from the HMDA/MM data likely reflect the non-prime focus of New Century. While the differences between the two data sets are not great, they are nevertheless driven by this self-selection, and thus our results should be considered with this in mind.

**Table 1**  
Summary statistics, Hispanic borrowers, by English proficiency area.

	All (1)	English proficient area (2)	Limited English proficient area (3)	T-Statistic
English proficient area	0.352 (0.478)	–	–	
Use Hispanic broker	0.698 (0.459)	0.671 (0.470)	0.747 (0.435)	11.470
Annual percentage rate	9.629 (1.354)	9.657 (1.347)	9.579 (1.366)	–3.977
Combined monthly income	7528.857 (3820.171)	7578.210 (3925.912)	7438.067 (3616.083)	–2.605
Married	0.527 (0.499)	0.535 (0.499)	0.513 (0.500)	–3.054
Self employed	0.341 (0.474)	0.326 (0.469)	0.368 (0.482)	6.142
Number of years on job	5.441 (4.897)	5.455 (5.021)	5.416 (4.661)	–0.564
Years in current residence	4.140 (4.242)	4.128 (4.324)	4.162 (4.089)	0.565
FICO score	654.935 (52.117)	652.334 (52.362)	659.719 (51.322)	9.902
Condo unit	0.079 (0.270)	0.079 (0.270)	0.079 (0.269)	0.000
2–4 units	0.095 (0.294)	0.077 (0.267)	0.129 (0.335)	12.274
Combined LTV (%)	82.222 (6.139)	82.247 (6.303)	82.175 (5.827)	–0.832
Loan amount (\$1000s)	251,416.600 (132,342.200)	246,728.800 (131,715.400)	260,040.600 (133,065.700)	6.958
Not primary residence	0.095 (0.293)	0.100 (0.300)	0.085 (0.279)	–3.550
No down payment	0.680 (0.466)	0.667 (0.471)	0.705 (0.456)	5.709
Prepayment penalty	0.829 (0.376)	0.823 (0.382)	0.841 (0.365)	3.362
Full documentation	0.361 (0.480)	0.386 (0.487)	0.314 (0.464)	–10.567
Fixed mortgage rate	0.065 (0.246)	0.069 (0.253)	0.057 (0.233)	–3.462
Number of observations	21,056	13,641	7415	

Notes: Sample based on Hispanic borrowers who purchased a 30-year, first-lien loan with non-missing observations for APR, broker surname, and borrower ethnicity. T-Statistic is based on the difference in means.

different, these differences do not tend to be quantitatively large. For instance, the difference in average FICO score is less than 7 points, although statistically significant.

### 3. Results

Table 2 presents the results of estimating Eq. (2). All standard errors are clustered at the ZIP code level to account for unobserved correlation. For brevity, we present only the coefficient estimates  $\hat{\beta}_2$ ,  $\hat{\beta}_3$  and  $\hat{\beta}_4$ . Full estimates are available upon request. The results in column (1) indicate that a mortgage originated by a Hispanic broker is associated with a higher price of 5.4 basis points for Hispanic borrowers. Hispanic borrowers that live in areas where Hispanics are primarily LEP receive on average, a higher price of 3.4 basis points. The coefficient of interest—the interaction term of the indicator variable for Hispanic borrowers that live in areas where Hispanics are primarily LEP and broker's Hispanic ethnicity—shows a lower APR of 5 basis points. This suggests that Hispanic borrowers living in areas where Hispanics are primarily LEP have a lower APR if they work with a Hispanic broker than if they had worked with a non-Hispanic white broker.

To assess whether we are measuring Hispanic borrower's English fluency rather than characteristics of the area, we also present in columns (2)–(4) results based on the population of non-Hispanic black, non-Hispanic Asian, and non-Hispanic white borrowers, respectively. Presumably, if the percentage of English fluency in the area proxies well the Hispanic borrower's English fluency, we would expect that  $\hat{\beta}_4 \neq 0$  for Hispanics only and  $\hat{\beta}_4 = 0$  for the set of non-Hispanic borrowers. As expected, the coefficient estimate is zero for the sample of non-Hispanic

black, non-Hispanic Asian, and non-Hispanic white borrowers. This gives us greater confidence that identifying areas where Hispanics are primarily LEP may be a good proxy for the Hispanic borrower's English

**Table 2**  
Effect of Hispanic broker on APR, by borrower's Hispanic ethnicity and race.

	Hispanic (1)	Non-Hispanic		
		Black (2)	Asian (3)	White (4)
Hispanic broker	0.054*** (0.013)	0.053* (0.027)	0.038** (0.041)	0.035* (0.016)
Less 50% speak	0.034+ (0.019)	0.037* (0.018)	0.056 (0.050)	0.015 (0.011)
Hispanic broker X less	–0.050* (0.024)	–0.037 (0.047)	–0.1 (0.084)	0.016 (0.032)
R2	0.779	0.771	0.787	0.793
Total number of observations	21,056	8080	1811	27,072

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

+ Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

fluency. Note that the coefficient estimate of  $\hat{\beta}_2$  is positive for a majority of the cases. This suggests that borrowers in general pay higher prices when working with Hispanic brokers, and there does not seem to be a net discount for Hispanic borrowers. Lastly, we note that borrowers that live in areas where Hispanics are primarily LEP pay higher prices as well, though it is only statistically significant for the Hispanic and non-Hispanic black population of borrowers. This is not surprising given that areas where Hispanics are primarily LEP is defined as the percent of Hispanics that do not speak English very well, rather than the area where the general population is predominantly defined as LEP.

Table 3 provides additional evidence that our estimate is capturing language affinity rather than other borrower or geographic characteristics.<sup>17</sup> Column (1) includes applications where the primary borrower and the co-borrower (if applicable) are Hispanic, while column (2) includes joint applications where at least one borrower is not Hispanic. Again, we find a significant effect on Hispanic borrowers that live in areas where Hispanics are primarily LEP, but no effect if there is at least one borrower that is not Hispanic. We note, however, that the number of observations for joint applications where one borrower is Hispanic and one borrower is not Hispanic is small and results may be driven by sample size issues. Nevertheless, these results are consistent with expectations; an additional borrower (presumably that is English proficient) may provide the language services needed if the Hispanic borrower has limited English proficiency. Finally, column (3) includes an additional robustness check of our classification of Hispanic brokers. Because our ethnicity proxy is based on surname, one might be concerned that the Hispanic broker variable includes female brokers who are married to a Hispanic, and therefore do not necessarily speak Spanish. While this would introduce measurement error and bias our estimates towards zero, we exclude female brokers in our analysis in column (3).<sup>18</sup> Results, however, are inconsistent with measurement error; the coefficient estimate in column (3) is not larger in absolute value nor is it significantly different from the coefficient estimate in column (1).

Lastly, we provide further evidence that our estimate is likely capturing language effects in Table 4. One would expect that language and communication skills are important in certain types of loans (e.g., home purchase<sup>19</sup>) and less so in others (e.g., refinance loans or second homes<sup>20</sup>). Borrowers that purchase their second homes or that refinance mortgages have experienced the mortgage process at least once when buying their primary home.<sup>21</sup> Additionally, lenders have additional information on the borrower's ability to repay based on their loan performance history. As such, both lenders and borrowers may not derive as much benefit from communicating in the same language other than the strict preference of speaking the same language. Assuming that the self-selection of borrowers into certain loan purposes or property types does not bias the estimates,<sup>22</sup> we would predict that for the sub-sample of Hispanic

**Table 3**  
Effect of Hispanic broker on APR for selected sample of borrowers.

	One or two Hispanic borrowers (1)	Joint Hispanic and non-Hispanic borrower (2)	Excluding female brokers (3)
Hispanic broker	0.051*** (0.013)	0.079** (0.066)	0.043*** (0.008)
Less 50% speak English well	0.037+ (0.019)	−0.060 (0.083)	0.027*** (0.008)
Hispanic broker X less 50% speak English well	−0.051* (0.024)	−0.077 (0.138)	−0.037* (0.016)
R2	0.778	0.842	0.77
Total number of observations	20,740	614	63,133

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, non-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

+ Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

borrowers who refinance their home or who purchase their second home, the magnitude of the coefficient of  $\hat{\beta}_4$  is lower or the estimate is not statistically significant. Columns (1) and (2) of Table 4 present the results of this analysis. For completeness, we present results solely for Hispanic borrowers that purchase a home mortgage for their primary residence (see column (3) of Table 4).

As expected, the effect of language is zero for the sub-sample of Hispanic borrowers that refinance their loan and for those that purchase their second home.<sup>23</sup> Interestingly, the coefficient of  $\hat{\beta}_2$  is positive in column (1) and similar in magnitude to that of the sub-sample of home purchase loans, but not significant in column (2). This suggests that there may some sorting across Hispanic borrowers and Hispanic brokers. The coefficient of  $\hat{\beta}_3$ , on the other hand, is effectively zero for both refinance and second home loans in contrast to home purchase loans and primary resident loans. Again, this may underscore the assumption that language barriers may not be as important once borrowers understand the mortgage process.

Columns (4), (5), and (6) of Table 4 present results based on the sub-sample of Hispanic borrowers that had a full documentation, no/limited, and exclusively no documentation loan, respectively.<sup>24</sup> While no documentation requires less documentation than full documentation loans, it also may require more communication between the broker and borrower to process the loan. As such, we would expect the magnitude of the coefficient of  $\hat{\beta}_4$  to be lower for those that had a full documentation loan compared to those that had a no documentation loan.

Again, results confirm our expectations; the coefficient of  $\hat{\beta}_4$  is zero for the sub-sample of full documentation loans (column (4)) while the effect of language affinity is negative and statistically significant for the sub-sample of no documentation loans (columns (5)–(6)). Moreover, the magnitude of  $\hat{\beta}_4$  is slightly larger than the main results in Table 2. Similar to results for the refinance sub-sample, the coefficient of  $\hat{\beta}_3$  is zero for full documentation loans, while it is positive for no/limited documentation

<sup>17</sup> We thank an anonymous referee in pointing out this concern.

<sup>18</sup> We thank an anonymous referee in pointing out this concern. We define whether the broker is female or male based on the social security database that lists the number of occurrences of female and male first names where the names have occurred at least five times. Names are based on the cohorts 1960–1980, and exclude any names where the gender can be either female or male. More information on the database can be found in <http://www.ssa.gov/oact/babynames/limits.html>.

<sup>19</sup> The data includes a variable that distinguishes different types of purposes of the loan—home purchase, home improvement, and refinance loans. Unless otherwise stated, our analyses are based on the subset of loans that are specifically home purchase loans.

<sup>20</sup> We also include investor homes as part of the classification of second homes. This may lead to measurement error if the borrower's investor home is the first home purchased.

<sup>21</sup> Borrowers that refinance are also potentially a different subset of borrowers because in order to refinance they need to have sufficiently good credit to be approved. If the threshold for qualification is stricter for those that refinance compared to those that purchase homes, then this is a source of potential bias in the quality of the refinance borrowers. Furthermore, to the extent that such borrowers rely on less language services due to unobservable factor than other Hispanic borrowers, this may also bias the results.

<sup>22</sup> We note that the bias from the self-selection of borrowers that refinance counteract any sorting of borrowers that purchase home loans. That is, Hispanic borrowers that live in areas where Hispanics are predominantly LEP are of lower creditworthiness when matched with a Hispanic broker than Hispanic borrowers that do not live in areas where Hispanics are predominantly LEP.

<sup>23</sup> We considered limiting the sample to Hispanics in areas where Hispanics are predominantly LEP and using the home-purchase sub-group as a control group and the refinance group as the treatment group to test this prediction. However, we did not test this because APR for refinance and home purchase mortgage products are not comparable and it is difficult to fully model the pricing differences between the two loan categories.

<sup>24</sup> Because of the sample size of limited documentation loans, we could not analyze limited documentation loans exclusively. Limited documentation suggests that the borrower only needs to provide some verification of the borrower's income and assets (i.e., W-2, bank statements, etc.). No documentation loans are loans for which the income and assets are not verified by the lender.

**Table 4**  
Effect of Hispanic broker on APR, by loan characteristics.

	Loan purpose	Owner occupancy		Documentation type		
	Refinance	Investor/2nd homes	Primary residence	Full	Limited doc/no doc	No doc
	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic broker	0.046** (0.017)	0.049 (0.040)	0.055*** (0.0135)	0.047* (0.021)	0.055*** (0.016)	0.054*** (0.016)
Less 50% speak English well	-0.001 (0.022)	0.027 (0.052)	0.0367+ (0.020)	0.006 (0.030)	0.047* (0.023)	0.046+ (0.024)
Hispanic broker X less 50% speak English well	-0.003 (0.027)	-0.012 (0.064)	-0.054* (0.025)	-0.022 (0.040)	-0.064* (0.027)	-0.067* (0.028)
R2	0.743	0.742	0.777	0.773	0.784	0.784
Total number of observations	19,472	1999	19,057	7596	13,460	13,000

Notes: The results are based on an ordinary least squares estimation of the APR. All standard errors are clustered at the ZIP code level. Robust standard errors are in parenthesis. Asterisks indicate statistical significance at the 10% level (+), 5% level (\*), or 1% level (\*\*), or 0.1% level (\*\*\*). The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

loans (columns (5)–(6)). This is consistent with results in the refinance sub-sample, where we interpret that different documentation types may reflect different needs and values of language affinity.

While the consistency of the aforementioned results makes us more confident that we are measuring the effect of language affinity, there may be still other explanations for the negative relationship of APR and Hispanic brokers that originate loans for Hispanic borrowers that live in areas where Hispanics are predominantly LEP. For example, greater broker competition in LEP areas or differential sorting of Hispanic borrowers in areas where Hispanics are LEP might explain our results. We assess the degree of supply driven and demand driven characteristics in turn in the subsequent sections.

### 3.1. Supply side

From the supply side, there may be two main reasons for lower APRs in areas where Hispanics are predominantly LEP that are not attributed to language or communication—increased market competition or lower marginal costs that are not associated with language services. Hispanic brokers, for example, may be highly concentrated in areas where Hispanics are predominantly LEP and, as such, the supply of Hispanic brokers may be greater than that of non-Hispanic white brokers. This may decrease the price to which they can charge Hispanic borrowers. Indeed, we find that the higher the percentage of Hispanics that are LEPs in an area, the greater the number of Hispanic and non-Hispanic white brokers.<sup>25</sup> We argue that while there is greater supply of brokers in areas where Hispanics are primarily LEP, we assume that brokers are mobile. That is, if Hispanic brokers were receiving lower prices because of the competitive nature in certain areas, Hispanic brokers may market themselves elsewhere where they can obtain greater profits. Additionally, because our results consistently show Hispanic brokers in general charge higher prices, we do not believe that greater market competition is driving our results.<sup>26</sup>

Nevertheless, to assess the effect of market competition, we include fixed effects at the county and MSA levels. At the most granular level, rate sheets are determined at the MSA or county level, though in general, they are determined at the state or census region level.<sup>27</sup> As such,

<sup>25</sup> As an additional analysis, we look at the number of non-Hispanic white brokers in a given area to calculate the correlation coefficient between the number of brokers and the percentage of Hispanics that are LEP. We find that the correlation coefficient is 0.19 overall and the correlation coefficient between the number of non-Hispanic white brokers and percentage of LEP Hispanics is 0.08. This suggests, that there is a greater number of brokers “targeting” LEP areas, but it is disproportionately more Hispanic brokers than non-Hispanic white brokers.

<sup>26</sup> The estimates of  $\beta_1$  in Table 2 are positive.

<sup>27</sup> A more granular level of rate sheets may impose fair lending risks, and there are general recommendations that institutions should be very wary of charging different prices in different geographic locations (see <http://www.treliant.com/Portals/0/Documents/Uploads/2011/05/ANewFairLendingCrackdown.pdf>).

including county or MSA level fixed effects would incorporate differences across areas that reflect differences in rate sheets or broker market competition. The results of exploring this possibility using MSA or county fixed effects are shown in Table 5. Again, we find that the effect of a Hispanic borrower living in areas where Hispanics are predominantly LEP is associated with a lower APR if the loan was originated by a Hispanic broker, regardless of which market conditions are considered. We note that the magnitude is comparable to that of our main results in Table 2.

Aside from the lower costs of language or communication skills that Hispanic brokers may bear with Hispanic borrowers that live in areas where Hispanics are predominantly LEP, Hispanic brokers may have lower negotiating skills or less experience that may manifest in lower loan prices. We argue that if this were true, we would expect that the effect of working with a Hispanic broker would be associated with a lower APR in general. As the results from Tables 2–5 show, however,  $\hat{\beta}_2$  is generally positive and statistically significant. One may alternatively argue that Hispanic brokers originating loans in areas where Hispanics are predominantly LEP are different than Hispanic brokers originating loans in areas where Hispanics are not predominantly LEP. As such, results are driven by differences in characteristics of Hispanic and non-Hispanic white brokers that vary across the two areas. While we do not have information on broker's experience level or other characteristics, we have information on the broker's total compensation. If Hispanic brokers are demonstratively “different” than non-Hispanic white brokers in areas where Hispanics are predominantly LEP, we would expect that their broker compensation would reflect such differences.<sup>28</sup> Results in Table 6, however, suggest that this is not the case. The results in Table 6 show that Hispanic brokers receive compensation that is 2–3% greater than that for non-Hispanic white brokers. Brokers that originate mortgages to Hispanic borrowers that live in areas where Hispanics are predominantly LEP receive compensation that is 2.4% higher than if brokers originate mortgages to Hispanic borrowers that do not live in areas where Hispanics are predominantly LEP. If the broker is Hispanic, the compensation is even higher. This also provides additional evidence that the lower APRs are not driven by market competition. We would expect lower broker profits in areas where Hispanics are predominantly LEP if this were the case.

### 3.2. Demand side

While it does not seem to be the case that supply-side factors explain our results, there may be differences in borrower characteristics that are driving the results. Hispanic borrowers in areas where Hispanics are predominantly LEP, for example, may sort differentially towards Hispanic and non-Hispanic brokers. More specifically, certain types of Hispanic

<sup>28</sup> Broker compensation is based on the log of total broker points and fees. The analysis is based on the same population as in the main results.

**Table 5**  
Effect of Hispanic broker on APR, market fixed effects.

	(1)	(2)
Hispanic broker	0.059*** (0.013) <sup>+</sup>	0.063*** (0.013)**
Less 50% speak English well	0.0665*** (0.019)	0.062*** (0.019)
Hispanic broker X less 50% speak English well	−0.052* (0.022)	−0.044* (0.021)
County fixed effects	Yes	
MSA fixed effects		Yes
R2	0.782	0.782
Total number of observations	20,089	21,056

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

<sup>+</sup> Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

borrowers that live in areas where Hispanics are predominantly LEP may select Hispanic brokers. To assess potential sorting effects, we estimate a linear probability model<sup>29</sup> where the dependent variable is equal to 1 if the broker is Hispanic and 0, otherwise, and all of the dependent variables from Eq. (2) are interacted with an indicator variable equal to 1 if Hispanics in the areas are predominantly LEP. Column (1) of Table 7 shows results without any interactions across year and states, while column (2) presents results with year interactions and column (3), results of the interactions of state and year with the dummy indicator.

The probability of choosing a Hispanic broker is significantly higher if the borrower lives in areas where Hispanics are predominantly LEP. This suggests that there is a demand for speaking the same language. However, it is unclear whether these respective borrowers are more or less creditworthy than the average. While it seems that borrowers with higher credit scores are more likely to choose Hispanic brokers, the borrowers are more cash-constrained. Lower income and higher debt-to-income ratios are associated with a greater probability of choosing a Hispanic broker relative to a non-Hispanic white broker. We also note that while the independent variables are statistically significant, the interactions of these variables with living in an area where Hispanics are predominantly LEP are not. Not surprisingly, the joint significance of the interactions is not statistically significant at the 5% level for the baseline specification. However, interacting the indicator variables with that of state and year indicators, we find that the interaction effects are statistically significant at the 5% level. This suggests that there is differential sorting across year and state.

To further investigate the potential of borrower sorting, we analyze the effect of APR by loan types and property type. That is, we reestimate Eq. (2) on the sub-sample of adjustable rate mortgage (ARM) and fixed-rate loans, as well as on the sub-sample of single family home,<sup>30</sup> condominium, and 2–4 unit homes. Results are reported in Table 8. We find that the effect is mainly driven by borrowers who chose fixed-rate loans and single family homes. This result is somewhat surprising given that the probability of choosing a Hispanic broker was higher for Hispanic borrowers that were more cash-constrained and that more Hispanic borrowers in areas where Hispanics are predominantly LEP apply for loans that are for 2–4 unit homes. One would

<sup>29</sup> Results are similar if we estimate a probit model.

<sup>30</sup> Single family units include single family homes, PUD attached, or PUD detached homes.

**Table 6**  
Effect of broker compensation.

	(1)	(2)	(3)
Hispanic broker	0.028* (0.009)**	0.017 <sup>+</sup> (0.009)***	0.179* (0.008)
Less 50% speak English well	0.024 <sup>+</sup> (0.013)	0.023 <sup>+</sup> (0.014)	0.023 <sup>+</sup> (0.013)
Hispanic broker X less 50% speak English well	0.029 <sup>+</sup> (0.016)	0.025 (0.015)	0.025 (0.015)
MSA fixed effects		Yes	
County fixed effects			Yes
R2	0.204	0.208	0.23
Observations	20,961	20,002	20,961

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

<sup>+</sup> Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

have expected that cash constraints would induce borrowers to choose an adjustable-rate loan rather than a fixed-rate loan since initial monthly payments are lower. Unfortunately, data limits our ability to further unravel the source behind this relationship.

However, to the extent that these results reflect borrower characteristics and explain our results, we are able to estimate whether there are differences in the probability of default across broker ethnicity and Hispanic borrowers that live in areas where Hispanics are predominantly LEP. Table 9 presents a linear probability model<sup>31</sup> of the probability of default. We define the probability of default as the probability that the loan was 60 days past due at any point from origination up to October 2011.<sup>32</sup> We do not find that the type of broker affects the probability of default for Hispanic borrowers in areas where Hispanics are predominantly LEP. This suggests that the creditworthiness of Hispanic borrowers that live in areas where Hispanics are predominantly LEP and select Hispanic brokers does not easily explain lower loan prices. Interestingly, we find that in the overall sample, Hispanic borrowers are more likely to default if they work with a Hispanic broker than a non-Hispanic white broker. However, this result is not consistent across all sub-samples.

The aforementioned results suggest that the types of loans rather than borrower characteristics may explain the differences. Whether these respective loans were chosen because of better communication between the broker and borrower, because of differential preferences of Hispanic borrowers in areas where Hispanics are predominantly LEP, or because of product choice availability by Hispanic brokers that work with Hispanic borrowers in areas where Hispanics are predominantly LEP, unfortunately, cannot be disentangled. Undocumented Hispanics in some areas of the country, for instance, use an Income Tax Identification Number (ITIN) to obtain the so-called "ITIN mortgages" (Diaz McConnell and Marcelli, 2007). Our data, unfortunately, cannot identify such loans. However, the limited studies on ITIN loans indicate that they required more careful and intensive underwriting process (justifying higher fees) and that their respective loan outperforms other mortgage products (Del Rio, 2010). The tougher underwriting standards for ITIN mortgages may give less flexibility to Hispanic brokers to give special discounts to undocumented immigrants. Even when undocumented immigrants used Hispanic brokers to obtain ITIN

<sup>31</sup> Results are similar if we estimate a probit model.

<sup>32</sup> The default probability is derived by merging the New Century data with the Core Logic data of loan performance for subprime and Alt-A mortgage loans.

**Table 7**  
Linear probability of choosing Hispanic broker.

	(1)	(2)	(3)
Less than 50% English speaking (LEP)	1.779** (0.653)	1.688** (0.657)	1.466* (0.718)
Debt to income	2.146*** (0.607)	2.235*** (0.609)	2.177*** (0.610)
LEP * debt to income	-0.908 (1.005)	-1.183 (1.020)	-1.244 (1.030)
Log combined loan to value	0.034 (0.045)	0.033 (0.045)	0.046 (0.045)
LEP * log combined loan to value	-0.028 (0.087)	-0.024 (0.087)	-0.052 (0.088)
Log loan amount	-0.015 (0.015)	-0.017 (0.016)	-0.018 (0.016)
LEP * log loan amount	-0.011 (0.024)	-0.006 (0.024)	0.002 (0.028)
Log (income)	-0.050** (0.016)	-0.047** (0.016)	-0.048** (0.016)
LEP * log income	-0.003 (0.026)	-0.011 (0.027)	-0.008 (0.027)
Log FICO	0.336*** (0.053)	0.332*** (0.053)	0.322*** (0.053)
LEP * FICO score	-0.215* (0.089)	-0.202* (0.089)	-0.190* (0.089)
Full documentation	-0.022* (0.009)	-0.021* (0.009)	-0.024* (0.009)
LEP * full documentation	-0.001 (0.015)	-0.004 (0.015)	0.007 (0.015)
Fixed rate	0.004 (0.015)	0.004 (0.015)	0.007 (0.015)
LEP * fixed	0.004 (0.027)	0.003 (0.027)	0.002 (0.028)
Self employed	0.028** (0.009)	0.029*** (0.009)	0.024** (0.009)
LEP * self employed	-0.027+ (0.014)	-0.030* (0.014)	-0.016 (0.014)
Job tenure	0.002* (0.001)	0.002* (0.001)	0.002* (0.001)
LEP * job tenure	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Housing tenure	0 (0.001)	0 (0.001)	0 (0.001)
LEP * house tenure	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
No down payment	0.006 (0.011)	0.007 (0.011)	0.002 (0.011)
LEP * no down payment	0.02 (0.018)	0.018 (0.018)	0.027 (0.019)
Married	-0.013+ (0.007)	-0.013+ (0.007)	-0.014+ (0.007)
LEP * married	0.006 (0.012)	0.007 (0.012)	0.008 (0.012)
Not primary house	-0.025+ (0.015)	-0.023 (0.015)	-0.029+ (0.015)
LEP * not primary house	0.002 (0.026)	-0.003 (0.026)	0.004 (0.027)
Condo	-0.026+ (0.014)	-0.026+ (0.014)	-0.032* (0.014)
LEP * condo	-0.053* (0.023)	-0.054* (0.023)	-0.031 (0.024)
Multi-unit	0.071*** (0.015)	0.070*** (0.015)	0.096*** (0.015)
LEP * multi-unit	0.017 (0.021)	0.018 (0.021)	-0.034 (0.023)
Year fixed effects	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Year & LEP interactions		Yes	Yes
State & LEP interactions			Yes
Adjusted R2	0.093	0.093	0.099
Observations	22,994	22,994	22,994

Notes: The results are based on an ordinary least squares estimation of the probability of choosing a Hispanic broker for a sub-sample of 30 year, first-lien mortgages by Hispanic borrowers. Robust standard errors are in parentheses.

+ Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

loans, Hispanic brokers needed to spend more time preparing and underwriting ITIN loans than other loans. This would lead to higher fees for ITIN mortgages relative to non-ITIN mortgages. While we are unable to control for the effect of ITIN loans in our results, the existing studies suggests that  $\hat{\beta}_4$  may be biased towards zero if there are a significant portion of ITIN mortgages in our data.

We do note, however, that the APR for low/no documentation loans and fixed loans, are in general, higher than full documentation loans.<sup>33</sup> This suggests that the lower APR is not reflective of the loan product, per se, but potentially unobservable characteristics associated with the respective loan products.

#### 4. Conclusions

Because the ability to successfully engage in financial services impacts the ability of individuals to save, borrow, increase wealth, or own a home, the impact that English proficiency has on an individual's ability to engage in financial services may be the foundation from which all other socioeconomic outcomes originate. Our paper provides some evidence that language barriers may be an important component in assessing the difficulties that non-English speakers face in bridging the gaps with the rest of society. Our results show that Hispanic borrowers that live in areas where Hispanics are predominantly LEP have lower APRs if their loans are originated by Hispanic brokers relative to non-Hispanic white brokers. The lower APR translates to a present-value benefit of about \$478 over 5 years and \$1697 over 30 years, for the average loan amount and APR. We interpret these results as a reflection of the language barriers that LEP borrowers face during the mortgage process. With Americans struggling to make ends meet, and house prices continuing to increase, these savings are nontrivial.

We note, moreover, that our estimates may serve as a lower bound on the true effect of language barriers since our results are based on the population that actually applied and received a mortgage. Furthermore, it is a lower bound if those that were discouraged by the language barriers are not included.

Our results also contribute to the cultural affinity literature where we see differential effects in mortgage pricing if members are of the same race or ethnicity. The existing literature had only focused on loan disposition. Further, our results suggest that language affinity is also important in explaining differential outcomes.

Our results are presented with the caveat that measurement error is inherent in our formulation of the borrower's English fluency and broker's Spanish fluency. Enhanced data collection that includes the language(s) used during the application process would further the researchers' ability to examine this issue. Future studies would be able to support or refute our finding that a market for language services is not only functioning, but also provides important economic benefits to individuals that generally have lower socio-economic outcomes as a consequence of lower English ability. Also, our results are based on the New Century data, a large subprime lender, known for questionable lending practices. Given that the broker's behaviors are not representative of lender behavior in general, our conclusions are limited to New Century broker practices. Lastly, our various empirical specifications show that our results are mainly driven by fixed-rate loans rather than adjustable-rate loans. To the extent that there are unobservable characteristics that are associated with fixed-rate loans originated by Hispanic brokers that explain the lower APRs for Hispanic borrowers in areas where Hispanics are predominantly LEP, our conclusions may be weakened. Extensions of this research can incorporate the population of discouraged individuals who do not seek a mortgage because of the language barriers as well as results based on a more diverse population.

<sup>33</sup> For example, the average APR for no documentation loans is 9.72 compared to 9.48 for full documentation loans.



**Table 8**  
Effect of Hispanic broker on APR, by loan type.

	ARM (1)	Fixed rate (2)	Condo (3)	2–4 units (4)	Family (5)
Hispanic broker	0.043*** (0.011)	0.029** (0.051)	−0.007+ (0.041)	−0.045 (0.051)	0.066*** (0.014)
Less 50% speak English well	0.028 (0.016)	0.025 (0.076)	0.020 (0.062)	−0.074 (0.066)	0.043* (0.021)
Hispanic broker * less 50% speak English well	−0.028 (0.019)	−0.255* (0.129)	0.077 (0.071)	0.047 (0.072)	−0.065* (0.026)
R2	0.845	0.636	0.791	0.761	0.781
Observations	19,356	1700	1665	2008	17,383

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

- + Indicates statistical significance at the 10% level.
- \* Indicates statistical significance at the 5% level.
- \*\* Indicates statistical significance at the 1% level.
- \*\*\* Indicates statistical significance at the 0.1% level.

**Table 9**  
Linear probability of default, 60 days past due.

	All (1)	No documentation (2)	ARM (3)	Fixed (4)
Hispanic broker	0.015+ (0.008)	−0.0103* (0.011)	0.126** (0.009)	0.025*** (0.032)
Less 50% speak English well	−0.004 (0.013)	−0.004 (0.0157)	−0.004 (0.013)	−0.017 (0.048)
Hispanic broker X less 50% speak English well	0.001 (0.014)	−0.001 (0.018)	0.001 (0.015)	0.017 (0.055)
R2	.0295	0.321	0.310	0.206
Observations	19,177	11,800	17,715	1462

Notes: The results are based on an ordinary least squares estimation of whether the loan was 60 days past due at any point from origination up to October 2011. Sample is based on a merged data set of New Century and Core Logic. The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, not-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

- + Indicates statistical significance at the 10% level.
- \* Indicates statistical significance at the 5% level.
- \*\* Indicates statistical significance at the 1% level.
- \*\*\* Indicates statistical significance at the 0.1% level.

**Appendix 1**

To ensure that our sample of loan in our analysis is representative, we compare our sample of loans to the rest of home purchase loans originated in the New Century data, all loans originated in the New Century data, and loans taken from a merged data set from the OCC Mortgage Metrics database and HMDA data for 2004 to 2006.<sup>34</sup> In contrast to the data from New Century, which was a

<sup>34</sup> We focus only on the merged database of the OCC Mortgage Metrics and HMDA data since HMDA data are able to identify the borrower's ethnicity while the OCC Mortgage Metrics data provide information on the borrower's creditworthiness. Hence, the merged data sets contain a significantly lower number of loans than the 23 million in the OCC Mortgage Metrics alone.

subprime lender, the OCC Mortgage Metrics database provides data on first-lien residential mortgages serviced by selected national banks. Between 2007 and 2008, this included more than 23 million first-mortgage loans, comprising more than 90% of all mortgages by national banks and approximately 40% of all outstanding mortgages. Hence, by comparing the characteristics of borrowers in the OCC Mortgage Metrics database to that of our sample, we can infer whether our conclusions can be potentially generalized to the rest of the population.

Appendix Table 1 presents the sample of 30-year, first-lien loans for Hispanic borrowers from the HMDA/MM data for 2004–2006. While there are differences between our sample from the New Century data, including FICO, loan amount and condo/multi-unit property type, these samples are still comparable. The differences between the two data are likely due to the fact that New Century focused on non-prime loans—91% of the loans in the New Century data are adjustable-rate mortgages loans (ARM) vs. 3% of HMDA/MM loans—and thus possibly lower qualified borrowers.

**Appendix Table 1**  
HMDA/MM 2004–2006, Hispanics.

	30 yrs, 1st lien	
	HP	All
FICO	697.31 −62.25	697.76 −62.51
LTV	0.82 −0.2	0.74 −0.25
DTI	38.6 −147.85	45.21 −275.03
Loan amount (\$1000s)	175.14 −108.09	182.87 −110.81
ARM	0.03 −0.17	0.03 −0.18
Interest at origination	6.32 −0.78	6.33 −0.82
Owner occupied	0.91 −0.28	0.91 −0.28
Condo or multi-unit	0.01 −0.11	0.01 −0.1
Observations	144,841	226,741

Notes: The number of observations reflects the total count for Hispanics in each column category, not the number of non-missing observations.

**Appendix Table 2**

Effect of Hispanic broker on APR, robustness checks.

	Hispanic broker cut-offs:		Comparison group
	90%	70%	Non-Hispanic broker
	(1)	(2)	(3)
Hispanic broker	0.044** (0.016)	0.062*** (0.011)	0.047*** (0.012)
Less 50% speak English well	0.012 (0.023)	0.037* (0.016)	0.028+ (0.017)
Hispanic broker + less 50% speak English well	-0.036 (0.028)	-0.056* (0.021)	-0.044* (0.021)
R2	0.778	0.781	0.787
Total number of observations	14,542	24,302	22,314

Notes: The results are based on an ordinary least squares estimation of the APR. Robust standard errors are in parentheses. The OLS regressions include a condominium indicator, multi-unit family house indicator, log of combined loan-to-value, log of loan amount, non-primary residence indicator, no-down-payment indicator, prepayment penalty indicator, full-documentation indicator, and fixed rate indicator, marital status, self-employed indicator, log of household income, number of years in present job, number of years in current residence, log of FICO score, indicators for the year of origination, and indicators for the property's state of residence.

+ Indicates statistical significance at the 10% level.

\* Indicates statistical significance at the 5% level.

\*\* Indicates statistical significance at the 1% level.

\*\*\* Indicates statistical significance at the 0.1% level.

## References

- Arrow, Kenneth, 1971. Some mathematical models of race discrimination in the labor market. In: Pascal, Anthony H. (Ed.), *Racial Discrimination in Economic Life*. Lexington Books, New York, pp. 83–102.
- Becker, Gary S., 1994. *The Economics of Discrimination*. The University of Chicago Press, Chicago, IL.
- Berndt, Antje, Hollifield, Burton, Sandas, Patrick, 2012. What Broker Charges Reveal About Mortgage Credit Risk. Carnegie Mellon University and University of Virginia.
- Bertrand, Marianne, Luttmer, Erzo F.P., Mullainathan, Sendhil, 2000. Network effects and welfare cultures. *Q. J. Econ.* 115 (3), 1019–1055.
- Borjas, George J., 1992. Ethnic capital and intergenerational mobility. *Q. J. Econ.* 107 (1), 123–150.
- Borjas, George J., 2002. Homeownership in the immigrant population. *J. Urban Econ.* 52 (3), 448–476.
- Bostic, Raphael W., 2003. A test of cultural affinity in home mortgage lending. *J. Financ. Serv. Res.* 23 (2), 89–112.
- Calomiris, Charles W., Kahn, Charles M., Longhofer, Stanely D., 1994. Housing-finance intervention and private incentives: helping minorities and the poor. *J. Money, Credit, Bank.* 26 (3), 634–674.
- Chiswick, Barry R., Miller, Paul W., 1995. The endogeneity between language and earnings: international analyses. *J. Labor Econ.* 13 (2), 246–288.
- Cortes, Alvaro, Herbert, Christopher E., Wilson, Erin, Clay, Elizabeth, 2006. *Improving Homeownership Opportunities for Hispanic Families: A Review of the Literature*. Housing and Urban Development, Washington, DC.
- Dávila, Alberto, Mora, Marie T., 2001. Hispanic ethnicity, English-skill investments, and earnings. *Ind. Relat.* 40 (1), 83–88.
- Del Rio, Deyanira, 2010. Mortgage Lending and Foreclosure in Immigrant Communities: Expanding Fair Housing and Fair Lending Opportunity Among Low Income and Undocumented Immigrants. Kirwan Institute, Ohio State University, Columbus, OH.
- Diaz McConnel, Eileen, Marcelli, Enrico A., 2007. Buying into the American Dream? Mexican immigrants, legal status, and homeownership in Los Angeles county. *Soc. Sci. Q.* 88 (1), 199–222.
- Elliot, Marc N., Morrison, Peter A., Freemont, Daniel F., McCaffrey, Daniel, Pantoja, Philip, Lurie, Nicole, 2009. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Serv. Outcomes Res. Methodol.* 9 (2), 69–83.
- Gonzalez, Arturo, 2013. The evolution of Hispanic literacy in the 21st century: from the first to the third generation. In: Mora, Marie T., Dávila, Alberto (Eds.), *The Economic Status of the Hispanic Population*. Information Age Publishing, Inc., Charlotte, NC.
- Haurin, Donald R., Rosenthal, Stuart S., 2009. Language, agglomeration and Hispanic homeownership. *Real Estate Econ.* 37 (2), 155–183.
- Hunter, William, Walker, Mary Beth, 1996. The cultural affinity hypothesis and mortgage lending decisions. *J. Real Estate Financ. Econ.* 13 (1), 57–70.
- Jones, Greg, 2012. Mortgage lending: confusing in every language. *Loyola Consum. Law Rev.* 24 (4), 661–680.
- Kochhar, Rakesh, Gonzalez-Barrera, Ana, Dockterman, Daniel, 2009. *Through Boom and Bust: Minorities, Immigrants and Homeownership*. Pew Hispanic Center, Washington, DC.
- Lang, Kevin, 1986. A language theory of discrimination. *Q. J. Econ.* 101 (2), 363–382.
- Lewbel, Arthur, 2007. Estimation of average treatment effects with misclassification. *Econometrica* 75 (2), 537–551.
- Longhofer, Stanely D., 1996. Cultural affinity and mortgage discrimination. *Fed. Reserv. Bank Clevel. Econ. Rev.* 32 (Third Quarter), 12–24.
- Mosaica, 2004. Focus group findings: cross-site report Prepared for the Congressional Hispanic Caucus Institute, Washington, DC.
- Ratner, Mitchell S., 1996. Many routes to homeownership: a four-site ethnographic study of minority and immigrant experiences. *Hous. Policy Debate* 7 (1), 103–145.
- Toussaint-Comeau, Maude, 2000. *Ethnic Immigrant Enclaves and Homeownership: A Case Study of an Urban Hispanic Community*. Federal Reserve Board of Chicago, Chicago, IL.
- Zavodny, Madeline, 2000. The Effects of Official English Laws on Limited-English-Proficient Workers. *J. Labor Econ.* University of Chicago Press 18 (3), 427–452.