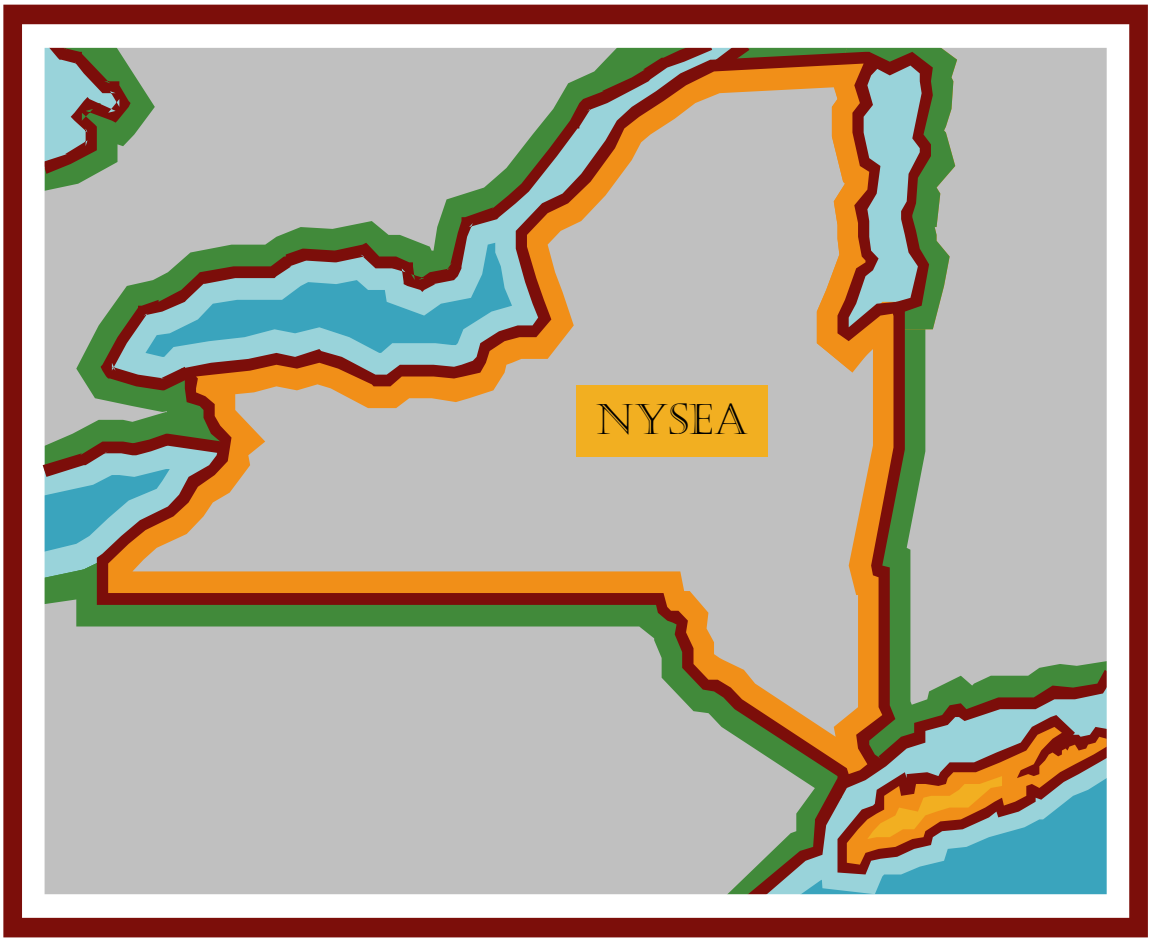
NEW YORK ECONOMIC REVIEW

FALL 2018



JOURNAL OF THE NEW YORK STATE ECONOMICS ASSOCIATION

VOLUME XLIX

NEW YORK STATE ECONOMICS ASSOCIATION

FOUNDED 1948

Vol. 49, Fall 2018

CONTENTS

ARTICLES

EDITORIAL

The *New York Economic Review* (NYER) is an annual publication of the New York State Economics Association (NYSEA). The NYER publishes theoretical and empirical articles, and also interpretive reviews of the literature in the fields of economics and finance. All well-written, original manuscripts are welcome for consideration at the NYER. We also encourage the submission of short articles and replication studies. Special Issue proposals are welcome and require a minimum of 4 papers to be included in the proposal, as well as a list of suggested referees.

MANUSCRIPT GUIDELINES

1. Please submit three copies of a manuscript.

2. All manuscripts are to be typed, double spaced and proofread. Prepared on a IBM PC/compatible computer in Microsoft Word format, the computer disk should be submitted in addition to the three hard copies.

3. All charts and graphs *must* be reproduction quality (Microsoft Word or Excel).

4. Footnotes should appear at the end of the article under the heading of "Endnotes."

5. Citations in the text should include the author and year of publication, as found in the references, in brackets. For instance (Marshall, 1980).

6. A compilation of bibliographic entries should appear at the very end of the manuscript under the heading "References."

Manuscript submissions should be emailed to editor David Vitt at david.vitt@farmingdale.edu

The NYER is cited in:

- EconLit by the American Economic Association
- Ulrich's International Periodicals Directory
- Cabell's Directory of Publishing Opportunities in Business and Economics

ISSN NO: 1090-5693

Do Audit Fees Influence Credit Risk and Asymmetric Information Problems? Evidence from the Syndicated Loan Market[1]

Lewis Gaul[*] Pinar Uysal[**]

ABSTRACT

The purpose of financial statement audits is to provide reasonable assurance that accounting records are free from material errors. We examine whether an increase in the demand for auditing services, which increases audit fees, is associated with a decrease in borrowers' credit risk and asymmetric information problems in the syndicated loan market. We assert that an increase in the quantity of auditing services purchased reduces the likelihood of an accounting error because auditors exert more effort verifying the accuracy of accounting records. We present empirical evidence that a demand-induced increase in audit fees is associated with syndicated loans with lower interest rate spreads and shorter maturity lengths, which we interpret as evidence for audit fee increases reduce credit-risk and asymmetric information problems.

INTRODUCTION

Several studies suggest that audited financial statements influence the terms and structure of syndicated loans. However, there is little or no research examining the influence of *potential* financial statement errors.[2,3] Previous research suggests that in the syndicated loan market, a greater likelihood of potential accounting errors increases asymmetric information problems and increases borrowers' credit risk (Graham, Li, and Qiu 2007). In addition, Dye (1993) and Simunic (1980) suggest that audit fees may be related to the likelihood of potential accounting errors.[4] In this paper, we examine whether an increase in audit fees as the result of an increase in the demand for auditing services is related to credit risk and asymmetric information problems in the syndicated loan market. We suggest that an increase in demand for auditing services should be associated with a greater quantity (hours billed) and/or price (hourly fee) of auditing services purchased, resulting in greater total audit fees.[5] We speculate that the extent of borrowers' credit risk and asymmetric information problems is decreasing in the quantity of auditing services purchased.[6] We use data on auditing fees and syndicated loans to provide evidence that an increase in auditing fees, due to an increase in the demand for auditing services, is negatively associated with the interest rates and the maturity length of syndicated loans.[7] In addition, we find that total audit fees are positively associated with the number of lenders in syndicated loans; however, we are unable to discern whether this result is attributable to supply of or demand for auditing services. We identify an increase in the demand for auditing services through an instrumental variable procedure with instruments that are expected

[*] Senior Economist, Office of the Comptroller of the Currency, Lewis.Gaul@occ.treas.gov

[**] Senior Economist, Federal Reserve Board of Governors, Pinar.Uysal@frb.gov

to shift the demand for auditing services rather than the supply of auditing services. We argue that our results are consistent with the argument that an increase in the demand for auditing services is associated with a decline in credit risk and asymmetric information problems in the syndicated loan market due to a decrease in the likelihood of potential accounting errors.

In the syndicated loan market, potential accounting errors exacerbate credit risk and asymmetric information problems. A greater likelihood of potential accounting errors reduces borrowers' expected future profitability thereby increasing borrowers' credit risk.[8] In addition, since borrowers are likely to have better information regarding their own characteristics compared to what would even be presented in the most accurate financial statements, potential accounting errors increase asymmetric information problems between borrowers and lenders. The Securities Exchange Act of 1934 mandates that all publicly traded firms have their financial statements audited by an independent external auditor. These mandatory audits potentially reduce firms' credit risk and asymmetric information problems by verifying that financial statements are accurate, in the sense that the statements adhere to generally accepted accounting principles (GAAP) and that users can be reasonably assured that financial statements are free from a material error.[9]

While all publicly traded firms are required to have their financial statements audited, firms have discretion to determine the quantity of auditing services purchased. In other words, firms are free to determine their individual demand for auditing services beyond a minimum quantity. Firms may demand a greater quantity of auditing services beyond the minimum amount to further decrease the likelihood of an accounting error, which should increase the accuracy of their financial statements.[10] Hence, an increase in auditing fees due to a shift in the demand for auditing services should be associated with a decline in firms' credit risk and asymmetric information problems in the syndicated loan market.[11]

Likewise, as firms are free to choose their own demand for auditing services, auditing firms individually choose their own supply of auditing services, which are determined by the costs of providing an audit. The costs of supplying auditing services include both the costs of physically performing audits and the expected future legal liabilities associated with audits. Auditing firms typically face legal liability from audits when a material accounting error that misleads investors is revealed, and auditors fail to detect the accounting error due to negligence in providing audits.[12] Since expected legal liabilities are derived from the expected likelihood of an accounting error or not detecting an accounting error, an increase in auditing fees due to a shift in the supply should be associated with an increase in credit risk and asymmetric information problems in the syndicated loan market. In addition, a shift in the supply could also result in a greater likelihood of an accounting error.[13]

In this paper, we examine whether an increase in audit fees paid by firms due to an increase in the demand for auditing services is related to credit risk and asymmetric information problems in the syndicated loan market. We base our examination, in part, on the theory that if an increase in auditing fees is due to an increase in the demand for auditing services, then an increase in audit fees is associated with a decrease in credit risk and asymmetric information problems. Our examination is also based on the previous theoretical

and empirical literature discussing the impact of credit risk and asymmetric information on debt contract terms, which predicts that borrowers with greater credit risk and asymmetric information problems receive loans: (1) with higher interest rates (Diamond 1984); (2) that are more difficult to sell (Leland and Pyle 1977; Diamond 1984; Holmstrom and Tirole 1997; Sufi 2007; Ivashina 2008). Combining data on audit fees from the Audit Analytics database and data on the price and non-price terms of syndicated loan contracts from the DealScan database, we test the hypothesis that if an increase in auditing fees is associated with an decrease in credit risk and asymmetric information, then: (1) the interest spread on a syndicated loan should be negatively associated with auditing fees; (2) the number of lenders in a syndicate should be positively associated with auditing fees; and (3) the maturity length of a syndicated loan should be negatively associated with auditing fees.

There are numerous complications associated with empirically testing our hypothesis. The primary complication is that audit fees are determined by the interaction of the supply and demand for auditing services. Hence, an increase in auditing fees may be associated with either an increase or decrease in the price and/or quantity of auditing services purchased.[14] In addition auditing fees may be endogenous to syndicated loan contract terms, correlated with unobserved and omitted control variables, and our proxy for auditing fees may be measured with error. In order to overcome these complications, we undertake our analysis with a generalized method of moments (GMM) instrumental variables estimator. We use instruments for the size of firms' inventory, accounts receivable, number of operating segments, and dispersion of economic activity among operating segments as instruments for auditing fees. We choose these instruments first of all because: (1) they are likely to be primary determinants of the demand for auditing fees; (2) they are not endogenous to loan contract terms; (3) they are uncorrelated with relevant unobservable or unintentionally omitted variables; (4) they are uncorrelated with measurement error in auditing fees; and (5) we perceive no strong theoretical argument as to why these variables should instead be used as explanatory variables for loan terms.

Our results indicate that audit fees are associated with syndicated loans with lower interest rates, shorter maturity lengths, and a greater number of lenders. However, our results are consistent with the assertion that demand-induced increases in audit fees influence the interest rate and maturity length of syndicated loans. Our results do not allow us to discern whether increases in the demand for auditing services influence the number of lenders. Overall, we interpret our results as supporting the proposition that an increase in auditing fees, due to a shift in the demand for auditing services, is associated with a decrease in credit risk and asymmetric information problems in the syndicated loan market.[15,16]

Our results are important for several reasons. First, the only study examining the implications of accounting errors or financial statement accuracy for debt contracting is Graham et al. (2007). However, these authors examine the implications of *realized* accounting errors for debt contracting. In contrast, our study examines the impact of *potential* accounting errors for debt contracting. These authors suggest that accounting errors increase credit risk as perceived by lenders because lenders usually lower expectations about borrowers' profitability following accounting errors, and accounting errors increase asymmetric

information problems because financial statement data is less reliable, which widens the information gap between borrowers and lenders. Moreover, our study is the first examining the empirical implications of audit fees for debt contracting.

The second reason our results are important is because the syndicated loan market is a primary source of financing for large publicly traded corporations, and our results provide additional insights regarding the impact of credit risk and asymmetric information in this market.

The third reason our results are important is that our results suggest auditing services mitigate asymmetric information problems with outside investors as intended by the Securities Exchange Act of 1934. Several observers have raised concerns that greater audit fees are no more than auditees paying to get away with accounting malfeasance, but our results suggest that lenders in the syndicated loan market associate a greater quantity of auditing services purchased with a decline in credit risk and asymmetric information problems.

Our results contribute to several literatures. First, our results contribute to the literature regarding the determinants of auditing fees. An implication of the theory by Simunic (1980) is that expected litigation costs are a primary determinant of auditing fees. Dye (1993) provides a model where the supply of auditing fees depends on expected litigation costs resulting from accounting errors, and the demand for auditing services depends on the benefits of more accurate financial statements. Carcello, Hermanson, Neal, and Riley (2002) present empirical evidence that audit fees are greater for better corporate boards, which implies the demand for more accurate accounting records is a determinant of auditing fees. We find that an increase in the demand for auditing services, which raises audit fees and the quantity of audit services purchased, is consistent with a reduction in credit risk and asymmetric information in the syndicated loan market. We interpret this result as suggesting that an increase in the quantity of auditing services purchased is associated with a reduced likelihood of financial statement errors.

Our results contribute to the literature regarding audit fee determination by providing additional evidence that an increase in audit fees, as a consequence of an increase in the demand for auditing services, is associated with a reduction in the likelihood of a financial statement error. Moreover, our results contribute to this literature by providing an econometric approach that attempts to separate the influence of the demand for auditing services on audit fees from the impact of the supply for auditing services.

Our results contribute to the literature regarding the influence of asymmetric information problems for the terms of syndicated loan contracts. Diamond (1984) presents a model where the cost of bank loan financing is increasing in the amount of resources lenders allocate to monitoring borrowers to overcome asymmetric information problems. Our results are consistent with the assertion that audit fees reduce asymmetric information, thereby reducing the amount of resources lenders must allocate to overcoming asymmetric information problems, therefore reducing borrowing costs.

Additionally, our results contribute to the literature regarding the incentives to produce information. Several studies provide theoretical justifications regarding barriers to information production about firms' creditworthiness. Hirshleifer (1971) suggests that agents producing information may have a hard time

credibly convincing other users that they have produced valuable information. Grossman and Stiglitz (1980) argue that it may not be economically rational to produce information if the producer cannot be certain that their information cannot be resold or transferred without their approval, thereby diminishing the returns to information production. Our results are consistent with the rationale that firms find it beneficial to pay for the production of information that can be used by anyone at zero cost, and that lenders in the syndicated loan market find this information credible.

THEORETICAL BACKGROUND

In the syndicated loan market, audited financial statements play a crucial role by influencing the extent of borrowers' credit risk and asymmetric information problems.[17] In a typical syndicated loan, a lead bank negotiates the non-price terms of a loan contract (loan amount, maturity length, collateral, covenants, performance pricing) with a borrower for an agreed-upon range of interest rates. Subsequently, the lead bank uses the negotiated loan contract terms to solicit a group of participant lenders willing to provide a portion of the loan's funding. Asymmetric information problems arise when borrowers have private information regarding their creditworthiness that they may use to the detriment of lenders' profitability. Borrowers' financial statements provide a noisy signal regarding borrowers' characteristics, reducing uncertainty regarding borrowers' creditworthiness, which mitigates asymmetric information problems.

Previous research examining the determinants of audit fees suggests that audit fees, which are defined as the price multiplied by the quantity of auditing services, may be either negatively or positively associated with the likelihood of an accounting error (Simunic 1980; Dye 1993), because an increase in auditing fees could be associated with either an increase or decrease in the quantity of auditing services purchased. Typically the quantity of auditing services is defined as hours worked by auditors and the price of auditing services is the hourly fee charged by auditors (Bell et al. 2001). These studies predict that an increase in the demand for auditing services should be associated with an increase in the price of auditing services, and an increase in the quantity of auditing services purchased. Any increase in the quantity of auditing services should be associated with a decline in the likelihood of an accounting error.

Typically, auditors verify accounting records by sampling a percentage of a unit of account. For example, when verifying the value of inventories or accounts receivables, auditors may not verify the value of each unit of inventory or every receivable but will instead verify the value of a percentage of inventories and receivables. Auditors may be able to verify a certain percentage of an account in a given number of hours, which provides a certain level of assurance that there are no accounting errors. Hence, if auditors sample a greater percentage of accounts, there should be an increase in the number of hours billed, and greater assurance that there are no errors in the valuation of these accounts.

The demand for auditing services depends on the benefits of more accurate financial statements, which include a decline in expected losses due to accounting errors and a reduction in asymmetric information problems between firms and outsiders (Graham et al. 2007). When accounting errors are realized, firms may be held liable and forced to pay damages to plaintiffs, which reduces their profitability. Profit expectations

are also reduced because accounting errors, more often than not, conceal unfavorable information regarding borrowers' future profitability. In addition, profitability also declines because firms often receive less favorable terms of trade in transactions following accounting errors, due to reputation damage caused by accounting errors. Potential accounting errors may exacerbate asymmetric information problems if borrowers have more knowledge regarding the correct information than lenders, and borrowers use this information advantage to the detriment of lenders' profitability.[18] Hence, an increase in audit fees due to an increase in the demand for auditing services should result in a decline in the likelihood that financial statement errors will be realized in the future, thereby reducing credit risk, and decreasing asymmetric information problems.

These models assert that the supply of auditing services is determined by the costs of physically providing an audit and the expected litigation costs associated with providing an audit. In terms of legal liability, auditors can be held individually liable if plaintiffs can prove that auditors did not provide audits consistent with generally accepted auditing standards (GAAS), and may suffer joint liability with audited firms' management if account records fail to adhere to generally accepted accounting principles (GAAP) (Dye 1993). Audits are generally considered to not comply with GAAS when audits do not adequately search for a material accounting error that misleads investors, and accounting records are considered to not comply with GAAP when there are material accounting errors that mislead investors. Since expected legal liabilities are derived from the likelihood of an accounting error, an increase in auditing fees as a result of a decrease in supply should be associated with an increase in credit risk and asymmetric information problems in the syndicated loan market. In addition, if a decrease in the supply for auditing services also lowers the quantity of auditing services purchased, this should also increase the likelihood of a potential accounting error.[19]

Given that potential accounting errors present an asymmetric information problem and additional credit risk, and that an increase in the demand for auditing fees should be related to a decrease in the likelihood of potential accounting errors, we can develop several empirical predictions regarding the association between audit fees and the terms of syndicated loan contracts, based on the literature discussing the impact of credit risk and asymmetric information on debt contracting.

The literature on loan contracting predicts that greater asymmetric information and credit risk is associated with higher loan interest rates. In the theories of Diamond (1991) and Boyd and Prescott (1986), lenders must exert more effort monitoring borrowers suffering from more severe asymmetric information problems, which raises the cost of loan financing. In addition, standard economic theory suggests that if a borrower's expected future profitability declines due to an accounting error, lenders will charge higher interest rates as compensation for greater default risk.

Several studies suggest that greater asymmetric information and credit risk should be associated with smaller lending syndicates. Bolton and Scharfstein (1996) present a model where lenders form smaller lending syndicates when default risk is greater in order to reduce bankruptcy costs, because it is easier to negotiate a resolution with fewer lenders. In addition, models by Leland and Pyle (1977), Diamond (1984), and Holmstrom and Tirole (1997) imply that lenders originating loans will retain a greater ownership stake in a loan to signal the quality of the loan and commit to monitoring the borrower. Sufi (2007) provides

empirical evidence that lead lenders in syndicates retain greater ownership stakes in syndicated loans, form smaller syndicates, and form more concentrated syndicates for borrowers suffering from more severe asymmetric information problems, particularly moral hazard problems. Ivashina (2008) presents evidence that lead lenders retain greater ownership stakes in loans in order to reduce asymmetric information problems.[20]

Two different studies provide empirical predictions regarding the impact of default risk and asymmetric information for the maturity of debt financing. Flannery (1986) presents a model with asymmetric information between borrowers and lenders where more creditworthy borrowers will issue short-term debt, when issuing debt requires the payment of transaction costs. They do so because paying repeated transaction costs to issue short-term debt, rather than issuing long-term debt, signals to credit markets that borrowers are more creditworthy. We suggest that greater auditing fees could be a transaction cost that firms face when issuing debt. Diamond (1991) presents a model where borrowers with both low and high credit quality will issue short-term debt and borrowers with moderate credit risk will issue long-term debt. Because our study focuses on borrowers with high or moderate levels of credit quality, we predict that borrowers with greater credit quality should borrow at shorter maturities.

Based upon the preceding discussion, we have three empirically testable predictions. If an increase in audit fees due to an increase in the demand for auditing services is associated with a decrease (increase) in credit risk and asymmetric information, then:

- audit fees are negatively (positively) associated with loan interest rates,
- audit fees are positively (negatively) associated with the number of lenders in a syndicated loan, and
- audit fees are negatively (positively) associated with the maturity of a syndicated loan.

EMPIRICAL MODEL AND SAMPLE SELECTION

We begin constructing our data sample with the Audit Analytics database, a database containing detailed audit information for more than 15,000 corporations filing public financial statements with the Securities and Exchange Commission (SEC). From this database, we gather data on firms' audit fees and non-audit fees. Audit fees include the cost of performing the audit, while non-audit fees include compensation for other ancillary services provided by auditors, such as tax preparation services.[21] We then merge the Audit Analytics database with the Loan Pricing Corporation's DealScan database, a database containing information regarding the price and non-price loan contract terms for loans to large corporations.[22] We combine observations from the merged Audit Analytics-DealScan database with accompanying financial statement data from Compustat and stock price data from the Center for Research in Securities Prices (CRSP) database. The unit of observation in our database is a loan facility obtained by a firm in a given fiscal year. The sample, with all loan facilities included, contains observations on 4,668 loan facilities merged to the aforementioned data sets and spans the years 2000-2007. We then randomly choose one loan facility

per year for each firm and arrive at a final sample of 2,971 loan facilities.[23] We are limited to this time span because the Audit Analytics database does not provide audit information prior to 2000. In addition, we note that all dependent variables constructed from the Audit Analytics, Compustat, and CRSP data are lagged one fiscal year prior to the beginning of the loan facility to ensure that the information was available to lenders when negotiating loan contract terms.

Our empirical exercise uses this data sample to estimate the following model:

$$Y_{i,t} = \delta\, fee_{i,t-1} + \beta' X + \omega_i + \gamma_t + \epsilon_{i,t} \quad (1)$$

Equation (1) presents the general model describing the interest rate spread, the number of lenders, and the maturity length of syndicated loans. The interest rate spread is the All-In-Drawn Spread from the DealScan database, which is the loan interest rate spread over LIBOR in basis points; the number of lenders is calculated as the log of number of lenders; and the maturity length of the loan is the log of the maturity length in days. In equation (1) the subscript *i* denotes the firm and the subscript *t* denotes the year. The dependent variable *Y* is the interest rate spread, the number of lenders, or the maturity length. The matrix *X* includes independent variables dated *t-1*, which serve as proxy for credit risk and asymmetric information problems, and are standard from the literature (Strahan 1999; Carey, Post, and Sharpe 1998; Hubbard, Kuttner, and Palia 2002; Graham et al. 2007). The error term is composed of three components: ω_i, which is the firm-specific error term; γ_t, the year-specific error term; and $\epsilon_{i,t}$, a white noise error term.

We calculate our proxy for audit fees as total audit fees plus non-audit fees divided by total assets. We use this measure to capture the possibility that firms compensate their auditors for their auditing activities by purchasing additional non-audit-related consulting services. For example, several studies suggest that firms may compensate auditing firms' for bearing additional litigation risk by purchasing additional services, such as tax preparation services.[24]

The observable risk characteristics in equation (1) that are included in *X* are intended to capture banks' pricing of risks related to credit risk and asymmetric information problems. These variables include: a proxy for the firm size (log of total assets), the leverage ratio (the book value of debt divided by the book value of assets), research and development (research and development expense divided by total assets), dividends (total dividends divided by total assets), current assets (current assets divided by total assets), the quick ratio (current assets minus current liabilities all divided by total liabilities), Tobin's average Q (the market value of equity plus the book value of debt divided by total assets), cumulative monthly stock returns from the previous fiscal year, and the standard deviation of monthly stock returns from the previous fiscal year.[25] In addition, we construct a proxy for the firms' Standard & Poor's (S&P) domestic issuer rating, which takes on 23 values, where the debt rating is more favorable for higher values of this indicator.[26] We expect control variables that capture greater (lesser) credit risk or asymmetric information problems to have the same (opposite) predicted associations with the dependent variables as audit fees. We expect greater values of the debt rating, total assets, current assets, the quick ratio, earnings before interest, taxes, depreciation, and

amortization (EBITDA), cumulative stock returns, and Tobin's average Q to be associated with less credit risk and asymmetric information problems; and we expect an increase in research and development spending, leverage, debt due in one year, and the standard deviation of stock returns to be associated with greater credit risk and asymmetric information problems. We offer no predictions as to how dividends should be associated with credit risk and asymmetric information problems.

The non-price loan terms capture how lenders use loan features to mitigate credit risk and asymmetric information problems (Strahan 1999). These include an indicator for whether or not the loan is secured, the log of the size of the loan facility, a dummy variable indicating whether the loan facility has financial covenants, a dummy variable indicating whether the loan facility has general covenants, and a dummy variable indicating whether a loan has performance pricing. In addition, we construct indicators for the loan type and purpose. Because we do not control for the endogeneity of loan contract terms, we do not offer any coefficient predictions. For example, Strahan (1999) finds that interest rate spreads are greater for secured loans, and Booth and Booth (2006) find that after controlling for the endogeneity of a loan being secured, secured loans carry lower interest rate spreads.

Our main objective is to obtain empirical estimates of the association between the quantity of auditing services and the dependent variables in equation (1). There are several complications to achieving this objective, which include: (1) we do not have data regarding the quantity of auditing services; (2) audit fees may be endogenous to the dependent variables; (3) audit fees may be measured with error; (4) audit fees may be correlated with unintentionally omitted or unobservable variables that explain the dependent variables. Therefore, we estimate equation (1) with instrumental variables generalized method of moments (IV-GMM) to identify the effect in audit fees, due to an increase in the demand for auditing services, on each dependent variable. IV-GMM parameter estimates are efficient and consistent in the presence of heteroskedasticity.

To implement the estimator, we need instrumental variables that are expected to be associated with an increase in demand for auditing services and not the supply of auditing services, correlated with audit fees, and uncorrelated with the error term in equation (1). We rely on the theory of Dye (1993) which implies that the demand for audit services depends on the benefits of more accurate accounting records, and the supply of audit fees is a function of the cost of performing an audit and the expected litigation costs associated with an audit's expected legal liability. Several studies suggest that audit liability is greatest when an audited firm defaults on a debt obligation, often leaving the auditing firm as the only entity with funds to reimburse creditors, which may suggest several proxies for default risk may be suitable instrumental variables that may capture a shift in the supply for audit services. However, since default risk should influence the demand for more accurate accounting records, these variables would likely be associated with a shift in the demand for auditing services. Hence, variables capturing credit risk would not be suitable instrumental variables for identifying changes in the supply and demand for auditing services. Therefore, we must choose another set of instrumental variables that are likely to be associated with a shift in the demand for auditing services. We derive our instrumental variables from Dye's implication that audit fees depend on the benefits of accurate

accounting records. Auditing clients wish to have an audit that provides a certain level of assurance that financial statements are free from errors.

We also consider the assertion of Bell et al (2001) and argue that the quantity of auditing services is captured by the hours billed by auditors and the price of auditing services is the hourly fee. Based on these assertions, it is reasonable to assume that auditing clients purchase a given amount of audit hours to achieve a certain level of assurance that accounting records are free from error. As previously mentioned, as variables that capture the marginal benefit of assurance are likely to be associated with credit quality or unobservable, we utilize variables that capture the need for clients to hire a greater number of auditing hours to achieve a given level of assurance. In a sense, these instrumental variables capture an increase in the quantity of auditor hours demanded, holding the marginal benefit of assurance constant. Our instruments include proxies for the scale of accounts receivable and inventories, the number of operating segments, and dispersion of economic activity among operating segments.

We justify accounts receivable and inventories based on the notion that auditors typically sample a certain percentage of these accounts to provide a given level of assurance. Hence, if a firm increases the scale of either of these items, holding all else constant, an audit would require a larger sample and a greater number of auditor hours, thus an increase in the quantity of auditing services purchased. We construct proxies for the scale of accounts receivable as total accounts receivable divided by total assets and total inventories divided by total assets. Instrumental variables are lagged to the fiscal year prior to the loan contract, concurrent with our audit fee proxy.

A second set of instrumental variables are based on the concept that more complex firms must purchase a greater number of audit hours to achieve a given level of assurance that accounting records are free from error. Our two proxies for complexity are the number of operating segments that comprise a firm and a Herfindahl Hirschmann Index (HHI) of sales among operating segments. If a firm has a greater number of segments, as stated by Simunic (1984), firms must have accounting records verified for more “decision centers.” In addition, if economic activity is more evenly dispersed among segments, then auditing activities will have to be dispersed among more decision centers.

In order to provide assurance that our instruments are appropriate, they must not: (1) shifts in the supply for auditing services (influence the marginal cost of providing an hour of auditing services), (2) not be endogenous to loan contract terms, (3) must not have measurement error correlated with the error term, (4) and must not be correlated with omitted variables. We argue it is reasonable to assume that our instrumental variables are robust to these potential problems. We argue that none of our variables influence the *marginal cost* of providing an *hour* of auditing services, but instead capture firms' increased demand for hours to achieve a given level of assurance that accounting records are free from a material error. We maintain that our instruments are not endogenous to loan contract terms because the instruments are dated as of the fiscal year prior to the loan contract. In addition, it is unlikely that measurement errors in our instruments are correlated with the error terms. Finally, we suggest that our variables are not correlated with any omitted variable because previous examinations of the empirical determinants of loan contract terms typically do not

include these “readily available” variables as explanatory variables for loan contract terms. Our host of other control variables, such as debt ratings and stock market valuations, likely better capture the information these variables may contain for loan contract terms.

ESTIMATION RESULTS

Before estimating equation (1) with instrumental variables with the IV-GMM method, we estimate the model simply by using Ordinary Least Squares (OLS). In OLS analysis we do not include instrumental variables for audit fees, however we control for industry and year effects and use robust standard errors clustered by firm. The results are presented in Table 1. In Column (1) we investigate the relation between all-in-drawn spread and the audit fees. The results indicate that audit fees have a positive and statistically significant association with loan interest rate spreads after controlling for firms' observable risk characteristics and non-price loan terms. Looking at the parameter estimates for the non-price loan terms in column (1), we see a negative and significant relation between the deal amount and the all-in-drawn spread. Similarly, the relation between number of lenders and the all-in-drawn spread is negative and significant. A secured loan is more likely to get a higher interest rate, whereas a loan that has performance pricing is more likely to get a lower interest rate. Higher-term loans, revolvers, and takeovers are also associated with higher spreads. Looking at the firm characteristics, we find that the lagged values of EBITDA, total assets, Tobin's average Q, sales, and firms' debt rating are negatively related to the All-In-Drawn Spread. An increase in leverage, cumulative stock returns and standard deviation of stock returns are positively related to the all-in-drawn spread. This result appears to indicate that audit fees are associated with greater credit risk and asymmetric information problems.

Next, in Column (3) we have the number of lenders as our dependent variable, and examine the effect of audit fees on the number of lenders. The coefficient on the audit fee is positive and statistically significant. The number of lenders is positively related with audit fees, consistent with the assertion that firms paying higher audit fees have lower asymmetric information and/or lower credit risk, and can borrow from syndicates that have more lenders. However, because the ordinary least squares (OLS) estimates do not identify a shift in either the demand or supply for auditing services, we cannot infer whether or not the increase in audit fees is associated with an increase in the quantity of auditing services. Examining the results for the observable risk characteristics, total assets are generally associated with fewer asymmetric information problems and are positively associated with the number of lenders. Interestingly, several observable risk characteristics have no significant association with the number of lenders. Firm sales are also positively related to the number of lenders, however the relation is negative for the variation in the stock returns and the number of lenders. The non-price loan terms have some explanatory power for the number of lenders. Revolver loans and the presence of general covenants and performance pricing is positively associated with the number of lenders, while secured loans, takeover and debt repayment loans have fewer lenders.

Table 1: OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	All-in-Drawn Spread		No. of Lenders		Maturity	
Audit Fee	26.961***	(6.563)	15.797**	(7.917)	-11.067**	(5.573)
Deal Amount	-0.069***	(0.019)	0.408***	(0.021)	0.054***	(0.018)
Maturity Length	-0.016	(0.027)				0.000
Number of Lenders	-0.045***	(0.016)				0.000
Secured/Unsecured	0.295***	(0.028)	-0.180***	(0.035)	0.032	(0.022)
Secured Dummy	-0.0701**	(0.028)	-0.0164	(0.038)	-0.0183	(0.026)
General Covenant Dummy	0.057	(0.040)	0.285***	(0.054)	0.085**	(0.037)
Financial Covenant Dummy	0.029	(0.032)	-0.026	(0.046)	-0.135***	(0.031)
Perf. Pricing Dummy	-0.078***	(0.029)	0.078*	(0.043)	0.047	(0.030)
Term Loan	0.466***	(0.057)	0.019	(0.056)	1.275***	(0.050)
Revolver/Line >= 1 Year	0.138***	(0.045)	0.101***	(0.038)	1.175***	(0.033)
Takeover	0.158***	(0.041)	-0.174**	(0.074)	-0.019	(0.046)
Debt Repay.	0.060*	(0.034)	-0.106*	(0.057)	0.101**	(0.040)
EBITDA	-0.840***	(0.138)	0.077	(0.137)	0.218	(0.135)
Total Assets	-0.042**	(0.016)	0.044**	(0.019)	-0.01	(0.015)
Tobin's Average Q	-0.088***	(0.011)	-0.018	(0.012)	-0.001	(0.009)
Sales	-0.025*	(0.014)	0.030*	(0.018)	0	(0.012)
Dividends	-0.914	(0.607)	-0.11	(0.749)	0.509	(0.400)
Leverage	0.359***	(0.069)	0.046	(0.092)	-0.012	(0.060)
Debt Due In One Year	0.002	(0.061)	-0.04	(0.068)	-0.049	(0.042)
Cum. Stock Returns	0.116**	(0.049)	0.057	(0.069)	0.092*	(0.053)
St. Dev. Stock Returns	0.767***	(0.188)	-1.415***	(0.219)	-0.199	(0.167)
Debt Rating	-0.134***	(0.006)	-0.01	(0.007)	-0.019***	(0.005)
Debt Rating Dummy	1.692***	(0.079)	0.218**	(0.098)	0.318***	(0.065)
Constant	6.161***	(0.318)	-6.635***	(0.327)	5.111***	(0.286)
R-squared	0.761		0.582		0.658	
N	2971		2971		2971	

Odd numbered columns report coefficient estimates, even numbered columns report p-value. * p < .10, ** p < .05, *** p < .01

In column (5), we find a negative and significant relation between audit fees and maturity length, which is consistent with Diamond (1991) and Flannery (1986) where borrowers with high credit quality will issue short-term debt. Audit fees may be similar to the transaction costs posited by Flannery (1986) or an

observable signal used to determine borrower's riskiness as suggested by Diamond (1991). The observable characteristics of the firms do not seem to matter very much for the determination of the maturity of the loan except for asset size and debt rating. However, the non-price loan terms have a substantial effect on the maturity. Higher term loans and revolvers, and loans that have general covenants are more likely to be made for a longer maturity. However, loans that have financial covenants are associated with shorter maturities.

Our main estimation results are presented in Table 2, where we use GMM two-step estimation with robust standard errors.[27] In this specification we use all the instruments; accounts receivable, inventories, number of operating segments, and HHI.[28] In column (1) the all-in-drawn spread is used as the dependent variable for the Equation (1). These results indicate that audit fees have a negative and statistically significant association with loan interest rate spreads. In addition, we note that audit fees retain significant explanatory power for loan spreads after controlling for firms' observable risk characteristics and non-price loan terms. This is important because more accurate financial statements may only influence loan terms through the weights banks place on information contained in the financial statements. However, our result implies that audit fees may be associated with a reduction in asymmetric information and credit risk beyond the more accurate information contained in financial statement data. Moreover, the results indicate that audit fees are negatively associated with loan spreads after controlling for credit risk, which could be a primary determinant of audit fees because audit fees are likely to depend on legal liabilities that ensue when borrowers are in financial distress. This result supports the assertion that, holding all else constant, more precise financial statements are associated with greater expected future profitability, hence, lower credit risk, and lower monitoring costs that need to be incurred to overcome asymmetric information problems.

Looking at the risk characteristics we see that firms' observable risk characteristics have explanatory power for loan spreads. Coefficient estimates are generally consistent with expectations. Loan interest rate spreads are decreasing in firms' total assets, Tobin's average Q, and EBITDA, which are all variables typically associated with lower credit risk and asymmetric information problems, while leverage and standard deviation of stock returns, generally associated with greater credit risk and asymmetric information problems, are positively associated with loan spreads. A better debt rating, which indicates a greater ability and willingness to repay debt, also reduces credit risk and asymmetric information problems, which results in lower interest rate spreads. For the non-price loan terms, an increase in the number of lenders and having performance pricing is negatively associated with loan spreads, while the presence of general or financial covenants are not significantly associated with loan spreads. Also, higher term loans, revolvers and takeovers seem to be positively related to the all-in-drawn spread. These results are all generally consistent with previous research examining the empirical determinants of loan spreads.

Next, in column (3) we replicate our analysis; however, this time we use number of lenders as our dependent variable. As can be seen in column (3), there is a positive but not significant relation between the number of lenders and audit fees. This means audit fees do not have any significant explanatory power. Looking at the firm characteristics, we see that higher total assets are positively associated with number of lenders, whereas higher Tobin's average Q, more variables stock returns or higher debt rating imply a higher

number of lenders. Looking at the non-price terms of the loans, we see that secured loans and takeovers can have fewer lenders, while loans with general covenants, performance pricing or revolver loans can have more lenders.

Table 2: GMM 2-Stage with All Instruments

	(1)	(2)	(3)	(4)	(5)	(6)
	All-in-Drawn Spread		**No. of Lenders**		**Maturity**	
Audit Fee	-63.619**	(31.964)	65.962	(42.172)	-45.875	(30.271)
Deal Amount	-0.078***	(0.018)	0.411***	(0.020)	0.052***	(0.017)
Maturity Length	-0.031	(0.027)				
Number of Lenders	-0.035**	(0.016)				
Secured/Unsecured	0.308***	(0.025)	-0.185***	(0.033)	0.035	(0.022)
Secured Dummy	-0.072***	(0.027)	-0.015	(0.036)	-0.021	(0.025)
General Covenant Dummy	0.049	(0.041)	0.285***	(0.053)	0.083**	(0.038)
Financial Covenant Dummy	0.034	(0.034)	-0.029	(0.044)	-0.130***	(0.030)
Perf. Pricing Dummy	-0.085***	(0.029)	0.084**	(0.042)	0.041	(0.029)
Term Loan	0.470***	(0.057)	0.025	(0.054)	1.267***	(0.047)
Revolver/Line >= 1 Year	0.131***	(0.044)	0.114***	(0.037)	1.167***	(0.032)
Takeover	0.174***	(0.043)	-0.181**	(0.072)	-0.014	(0.046)
Debt Repay.	0.0521	(0.036)	-0.102*	(0.058)	0.098**	(0.040)
EBITDA	-1.017***	(0.173)	0.178	(0.161)	0.138	(0.150)
Total Assets	-0.098***	(0.024)	0.076**	(0.031)	-0.032	(0.024)
Tobin's Average Q	-0.072***	(0.012)	-0.026**	(0.013)	0.005	(0.010)
Sales	-0.005	(0.014)	0.02	(0.019)	0.007	(0.012)
Dividends	-0.516	(0.576)	-0.348	(0.716)	0.624	(0.418)
Leverage	0.387***	(0.061)	0.032	(0.083)	0.005	(0.059)
Debt Due In One Year	0.053	(0.061)	-0.069	(0.071)	-0.026	(0.047)
Cum. Stock Returns	0.141***	(0.052)	0.042	(0.067)	0.088*	(0.052)
St. Dev. Stock Returns	0.975***	(0.209)	-1.533***	(0.225)	-0.142	(0.174)
Debt Rating	-0.128***	(0.006)	-0.014**	(0.007)	-0.017***	(0.005)
Debt Rating Dummy	1.612***	(0.075)	0.270***	(0.095)	0.294***	(0.063)
Constant	6.955***	(0.406)	-7.093***	(0.436)	5.412***	(0.350)
Hansen's J p-value	0.865		0.616		0.247	
Endogeneity	0.002		0.225		0.231	

Odd numbered columns report coefficient estimates, even numbered columns report p-value. * p < .10, ** p < .05, *** p < .01. Under-identification test p-values all less than 0.05. All models have N=2971 observations.

Afterward in column (5) we replicate our analysis by using maturity of the loan as our dependent variable. We find a negative but not significant association between audit fees and the maturity. Most of the characteristics of the firms do not seem to matter for the maturity. Higher cumulative stocks returns imply a longer maturity but higher debt rating implies a shorter maturity. Term loans and revolver loans are

associated with longer maturities. Also, if a loan has general covenants the maturity is longer, but if a loan has financial covenants the maturity is shorter.

It is important to note that identification of the exogenous influence of audit fees on the dependent variables relies on our instrumental variables being correlated with audit fees and uncorrelated with the error terms in equation (1). In other words, our instrumental variables must have reasonably potent correlation with audit fees in order to identify the exogenous influence of auditing fees on the dependent variables. Therefore, we examine the p-values from the test of under-identification to examine whether audit fees are reasonably correlated with our instrumental variables excluded from the second stage regression. The null hypothesis in the test of under-identification is that the instrumental variables excluded from the first stage are not correlated with audit fees. We reject the null hypothesis at the 1 percent level for all estimations in three columns. In order to make inferences as to the possible correlation between the instrumental variables and the error terms in equation (1), we examine the p-values for the test of over-identification. For the Hansen's J test of over-identification, the null hypothesis is that the instruments are uncorrelated with the error term, and we fail to reject the null hypothesis at any reasonable level for all three columns. In summary, we find no strong evidence rejecting the validity of our instrumental variables.

We also implement an endogeneity test, where under the null hypothesis audit fees can be treated as exogenous. For the estimation in column (1), we reject the null hypothesis, which implies that audit fees are endogenous to the all-in-drawn spread, and using an instrumental variables approach is relevant. However, checking the endogeneity test results in columns (3) and (5), we fail to reject the null hypothesis that Audit Fees are exogenous to the number of lenders and maturity. Since specification tests for the models presented in columns (3) and (5) reject the hypothesis that audit fees are exogenous to interest rate spreads, we do not elaborate on the meaning of the parameter estimates in these columns.

As can be seen from the first stage estimations are presented in Table A.2 in Appendix A, there is not a significant association between inventories and audit fees, and the number of segments and accounts receivable seem to matter the most among other instruments. Hence we replicate our analysis by using only number of segments and accounts receivable as instruments. The results are presented in Table 3.

The results in column (1) are in line with our previous findings that there is a negative and significant relationship between audit fees and all-in-drawn spread. The significance of the variables for the risk characteristics of the firms and the non-price terms of the loans do not change. The results in columns (3) and (5) are substantially different from our previous table. In column (3) where we have the number of lenders as our dependent variable, and the coefficient on the audit fee is positive and statistically significant. The number of lenders is positively related with the audit fee, consistent with the assertion that firms that pay higher audit fees have less asymmetric information and/or lower credit risk, and can borrow from syndicates that have more lenders. In column (5), where we have the number of lenders as our dependent variable, we find a negative and significant relation between the audit fee and the maturity length.

Next, we check for the under-identification test, and reject the null hypothesis that the instrumental variables excluded from the first stage are not correlated with audit fees for all three columns. For the

Hansen's J test of over-identification, we fail to reject the null hypothesis that the instruments are uncorrelated with the error term. Checking for the endogeneity, we find that audit fees cannot be treated as exogenous to the all-in-drawn spread as can be seen in column (1). However, we fail to reject the null hypothesis that the audit fees are exogenous to the number of lenders and the maturity as can be seen in columns (3) and (5).

Table 3: GMM 2-Stage with Accounts Receivables and No. of Segments

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	All-in-Drawn Spread		**No. of Lenders**		**Maturity**	
Audit Fee	-63.567*	(33.009)	78.223*	(43.558)	-57.361*	(31.256)
Deal Amount	-0.078***	(0.018)	0.411***	(0.020)	0.050***	(0.017)
Maturity Length	-0.031	(0.027)				
Number of Lenders	-0.035**	(0.016)				
Secured/Unsecured	0.308***	(0.025)	-0.187***	(0.033)	0.0367*	(0.022)
Secured Dummy	-0.072***	(0.027)	-0.014	(0.036)	-0.021	(0.025)
General Covenant Dummy	0.049	(0.041)	0.290***	(0.053)	0.084**	(0.039)
Financial Covenant Dummy	0.034	(0.034)	-0.031	(0.044)	-0.133***	(0.030)
Perf. Pricing Dummy	-0.085***	(0.029)	0.084**	(0.042)	0.041	(0.029)
Term Loan	0.470***	(0.057)	0.029	(0.054)	1.269***	(0.047)
Revolver/Line >= 1 Year	0.131***	(0.044)	0.117***	(0.037)	1.165***	(0.032)
Takeover	0.174***	(0.043)	-0.182**	(0.072)	-0.01	(0.047)
Debt Repay.	0.052	(0.036)	-0.099*	(0.058)	0.095**	(0.040)
EBITDA	-1.0171***	(0.173)	0.213	(0.165)	0.135	(0.151)
Total Assets	-0.098***	(0.025)	0.083***	(0.031)	-0.038	(0.024)
Tobin's Average Q	-0.072***	(0.012)	-0.029**	(0.013)	0.007	(0.011)
Sales	-0.005	(0.014)	0.016	(0.020)	0.011	(0.013)
Dividends	-0.517	(0.579)	-0.379	(0.719)	0.709*	(0.429)
Leverage	0.387***	(0.061)	0.029	(0.083)	0.009	(0.060)
Debt Due In One Year	0.053	(0.061)	-0.076	(0.071)	-0.022	(0.048)
Cum. Stock Returns	0.141***	(0.052)	0.04	(0.067)	0.102*	(0.053)
St. Dev. Stock Returns	0.974***	(0.209)	-1.560***	(0.227)	-0.107	(0.176)
Debt Rating	-0.128***	(0.006)	-0.014**	(0.007)	-0.016***	(0.005)
Debt Rating Dummy	1.612***	(0.076)	0.274***	(0.096)	0.277***	(0.065)
Constant	6.914***	(0.413)	-7.161***	(0.445)	5.481***	(0.353)
Hansen's J p-value	0.693		0.726		0.487	
Endogeneity	0.003		0.141		0.122	
N	2971		2971		2971	

Odd numbered columns report coefficient estimates, even numbered columns report p-value. * p < .10, ** p < .05, *** p < .01. Under-identification test p-values all less than 0.05.

So, we argue that since audit fees can be treated as exogenous, the estimation results in Table 1 where we do not employ instrumental variables is relevant, and there is a positive association between audit fees and number of lenders, and there is a negative association between audit fees and maturity, which is consistent with the argument that an exogenous increase in audit fees is associated with an decrease in credit risk and asymmetric information problems. However, as mentioned earlier, the results do not allow us to discern a specific interpretation of these results. Moreover, when comparing these results to those in Table 2 or Table 3, it may seem that we should assert our results provide mixed inferences regarding whether un-instrumented audit fees are associated with an increase or decrease in credit risk and asymmetric information problems. However, we remind the reader that we reject the null hypothesis that audit fees are exogenous to the interest rate spread, but not the number of lenders or the maturity.

CONCLUSION

Our estimation results support the view that increases in audit fees are driven by audit clients' demand for more accurate financial statements, which implies a lower likelihood of material accounting errors, resulting in a decline in credit risk and asymmetric information problems in the syndicated loan market. In addition, our results are consistent with the rationale that firms find it beneficial to pay for the production of information that can be used by anyone at zero cost, and that lenders in the syndicated loan market find this information credible. We interpret our results as suggesting that borrowers find it economically advantageous to substitute banks' monitoring with information production by auditing firms. Possibly this information could also be used by many financial market participants other than banks. In addition, our results suggest that audits serve the purpose stated in the Securities Exchange Act of 1934: that audits are intended to mitigate asymmetric information problems in financial markets.

ENDNOTES

1. We are grateful to Fabio Schiantarelli and Christopher Baum for thoughtful suggestions and comments. We thank the anonymous referee, participants of the workshop on "Challenges in Banking Research" organized by the Research Task Force of the Basel Committee on Banking Supervision. We also thank Lily Chin for editorial Assistance. The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Office of the Comptroller of the Currency, the U.S. Department of the Treasury, the Board of Governors of the Federal Reserve, or the Federal Reserve System.

2. A syndicated loan is a loan where two or more lenders provide a loan to a borrower. Typically, a lead lender negotiates the terms of a loan contract directly with a borrower for an agreed-upon range of interest rates. The lead lender then uses the negotiated terms of the loan contract to solicit participant lenders to provide a portion of the loan's funding. Usually, a lead lender provides funding for the residual portion of the loan that remains after soliciting financing from participants. Lead lenders typically transfer as much ownership of loans to participants as possible.

3. Recent research describing how audited financial statements potentially reduce asymmetric information problems in the syndicated loan market include Simons (1993), Preece and Mullineaux (1996), Dennis and Mullineaux (2000), Jones, Lang, and Nigro (2000), Lee and Mullineaux (2004), Panyagometh and Roberts (2002), Esty and Megginson (2003), and Sufi (2007).

4. For the remainder of the paper we refer to audit fees as the price multiplied by the quantity of auditing services. When needed, we refer to the price and quantity of auditing services separately. Consistent with Bell, Landsman, and Shackelford (2001), we interpret the unit of audit quantity as hours of auditing services, and the price of auditing services as the fee per hour of auditing services.

5. We define the quantity of auditing services as the number of hours billed by auditors and price of auditing services as the hourly fee charged by auditors.

6. Auditors reduce credit risk and asymmetric information problems by providing a certain level of assurance that financial records are free from a material error. We suggest that the level of assurance auditors provide is an increasing function of the number of auditing hours billed.

7. While it may seem controversial that firms with greater credit quality and fewer asymmetric information problems would borrow at shorter maturities, several empirical papers find similar results with similar data sets. For example, Strahan (1999) finds that borrowers with speculative grade debt ratings borrow at higher interest rates and longer maturities in the syndicated loan market. Moreover, Graham et al. (2007) examines the impact of realized accounting errors and finds firms with realized accounting errors borrow at shorter maturities because they are of intermediate to high risk. As we more fully articulate later on, we predict that the majority of our borrowers in our data sample are of moderate to high credit quality, which we will assert allows us to predict maturity is negatively associated with credit quality and asymmetric information problems. In addition, Graham et al. (2007) mentions the same point we make later on, that maturity is a non-monotonic function of credit quality. We mention this all early on, to be proactive in justifying this possibly counterintuitive prediction to readers.
8. Graham et al. (2007) suggests that when accounting errors are realized, firms' expected future profitability can decline for three reasons: (1) usually revealing accounting errors unveils unfavorable information regarding firms' expected future profitability; (2) firms typically have to pay damages to investors that have been misled by accounting errors; (3) the terms of trade firms face usually change unfavorably after the revelation of an accounting error. For example, customers of a manufacturing firm may believe that the firm may be more likely to declare bankruptcy in the future after an accounting error and customers may be less likely to purchase goods with warranties thinking that the firm will not remain in business to service the warranty.
9. A material error is an error that would change a decision made by a user of the financial statements.
10. For example, when verifying the value of a firm's assets, auditors may value a sample percentage of assets, rather all of the assets; however, firms can request that auditors sample a larger percentage of assets to increase confidence in the valuation of assets stated in their public financial statements.
11. As will be explained later, an increase in the demand for auditing services would imply an increase in the quantity of auditing services purchased which could reduce credit risk and asymmetric information problems.
12. Auditing firms should be held liable whenever there is a misleading accounting error that was left undetected because of negligently provided audits. Practically, errors are commonly found when borrowers are in financial distress. In addition, auditors are often found liable when auditing clients are in financial distress even when audits were not negligently provided.
13. Two mechanisms are at work here: (1) a decrease in supply is associated directly with expected legal liabilities, which should be associated with greater credit risk; (2) a decrease in the supply of auditing services is directly associated with a decrease in the quantity of auditing services purchased assuming a downward-sloping demand curve for auditing services.
14. As we discuss later, we assume the quantity of auditing services is the number of hours billed, and the price of auditing services is the hourly fee as in Bell, Landsman, and Shackelford (2001).
15. While investment in inventories could be determined, in part, by the cost of syndicated loan financing or other loan contract terms, our instrumental variables are lagged one period, which reduces the likelihood that the instrumental variables are endogenous to loan features.
16. We argue that our instrumental variables do not belong in the regression because any information these variables contain for credit risk is likely spanned by the other control variables and there are no definitive predictions regarding these variables and credit risk. For example, while inventories could reflect the existence of more collateral available in the event of default, an increase in inventories could also reflect an unexpected decline in sales.
17. See Simons 1993; Preece and Mullineaux 1996; Dennis and Mullineaux 2000; Jones et al. 2001; Lee and Mullineaux 2004; Panyagometh and Roberts 2002; Esty and Megginson 2003; Sufi 2007.
18. For example, borrowers may fraudulently misrepresent financial statement data in order to inflate financial markets' expectations of their future earnings, thereby distorting financial markets' perception of their credit risk. Hence, financial markets realize that financial statements may not reflect borrowers' true risk characteristics, exacerbating asymmetric information problems.
19. Audit fees could also increase due to an increase in the supply of auditing services if the price elasticity of demand is greater than one. While possible, we view this as a less likely scenario.
20. When we refer to asymmetric information problems in the syndicated loan market, we refer to two separate problems: those between lenders and borrowers, and those between lead lenders and participant lenders. We suggest that potential accounting errors influence these asymmetric information problems in the same manner.
21. As we later mention, we implement other measures of audit fees and the qualitative results are unchanged.
22. We gather the following loan information from the DealScan database: the loan interest rate, the number of lenders, the loan amount, whether the loan is secured or unsecured, whether the loan has financial or general covenants, whether the loan has performance pricing, the type of loan (i.e., loan commitment, term loan), the loan purpose, the loan seniority, and the distribution method (i.e., syndicated loan, sole lender loan).
23. Previous research follows this approach, for example, see Sufi (2007) and Ivashina (2008).
24. We also estimated all models using only audit-related fees divided by lagged total assets as our proxy for audit fees, and results were qualitatively similar.
25. Other studies examining the empirical determinants of loan contact terms include Carey, Post, and Sharpe (1998), Hubbard, Kuttner, and Palia (2002), Guner (2006) and Qian and Strahan (2007).
26. We set missing values of the debt rating equal to zero and generate an indicator variable equal to 1 when the debt rating is not missing.
27. For the two step estimations we used the Stata's routine by Baum, Schaffer, and Stillman (2008).
28. The first stage estimations are presented in Table A.2 in Appendix A.

REFERENCES

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. 2008. “IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation." July.

Bell, Timothy B., Wayne R. Landsman, and Douglas A. Shackelford. 2001. “Auditors Perceived Business Risk and Audit Fees: Analysis and Evidence." *Journal of Accounting Research*, 39 (June): 35-43.

Bolton, Patrick, and David S. Scharfstein. 1996. “Optimal Debt Structure and the Number of Creditors." *Journal of Political Economy*, 104 (1): 1-25 (February).

Booth, James R., and Lena Chua Booth. 2006. “Loan Collateral Decisions and Corporate Borrowing Costs." *Journal of Money Credit and Banking*, 38 (1): 67-90.

Boyd, John H., and Edward C. Prescott. 1986. “Financial intermediary-coalitions." *Journal of Economic Theory*, 38 (2): 211 - 232.

Carcello, Joseph V., Dana R. Hermanson, Terry L. Neal, and Richard A. Jr. Riley. 2002. “Board Characteristics and Audit Fees." *Contemporary Accounting Research*, 19 (3): 365-384.

Carey, Mark, Mitch Post, and Steven A. Sharpe. 1998. “Does Corporate Lending by Banks and Financial Companies Differ? Evidence on Specialization in Private Debt Contracting." *The Journal of Finance*, 53 (3): 845-878.

Dennis, Steven, and Donald Mullineaux. 2000. “Syndicated Loans." *Journal of Financial Intermediation*, 9:404-426.

Diamond, Douglas W. 1984. “Financial Intermediation and Delegated Monitoring." *Review of Economic Studies*, 51:393-414.

Diamond, Douglas W. 1991. “Monitoring and Reputation: The Choice Between Bank Loans and Privately-Placed Debt." *Journal of Political Economy*, 99:689-721.

Dye, Ronald A. 1993. “Auditing Standards, Legal Liability, and Auditor Wealth." *Journal of Political Economy*, 101 (5): 887-914.

Esty, Benjamin C., and William L. Megginson. 2003. “Creditor Rights, Enforcement and Debt Ownership Structure." *Journal of Financial and Quantitative Analysis*, 38:37-59.

Flannery, Mark J. 1986. “Asymmetric Information and Risky Debt Maturity Choice." *The Journal of Finance*, 41:18-38.

Graham, John R., Si Li, and Jiaping Qiu. 2007. “Corporate Misreporting and Bank Loan Contracting." NBER Working Paper, no. 13708.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. “On the Impossibility of Informationally Efficient Markets." *American Economic Review*, 70 (3): 393-408.

Guner, A. Burak. 2006. “Loan Sales and the Cost of Corporate Borrowing." *Review of Financial Studies*, 19 (2): 687-716.

Hirshleifer, Jack. 1971. “The Private and Social Value of Information and the Reward to Inventive Activity." *The Journal of Finance*, 61 (4): 561-574 (September).

Holmstrom, Bengt, and Jean Tirole. 1997. “Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics*, 112:663-691.

Hubbard, R. Glenn, Kenneth N. Kuttner, and Darius N. Palia. 2002. “Are There Bank Effects in Borrowers' Costs of Funds?: Evidence from a Matched Sample of Borrowers and Banks." *Journal of Business*, 75 (4): 559-581.

Ivashina, Victoria. 2008. “Asymmetric Information Effects on Syndicated Loan Rates." *Journal of Financial Economics*, 92: 300-319.

Jones, Jonathan D., William W. Lang, and Peter Nigro. 2000. “Recent Trends in Bank Loan Syndications: Evidence for 1995 to 1999." Office of the Comptroller of the Currency Working Paper.

Jones, Jonathan D., William W. Lang, Peter Nigro, and Richard A. Riley Jr. 2001. “Recent Trends in Bank Loan Syndications: Evidence for 1995-1999." Federal Reserve Bank of Chicago Proceedings, pp. 328-352.

Lee, Sang Whi, and Donald J. Mullineaux. 2004. “Monitoring, Financial Distress, and the Structure of Commercial Lending Syndicates." *Financial Management*, 33, no. 3.

Leland, Hayne E., and David H. Pyle. 1977. “Informational Asymmetries, Financial Structure, and Financial Intermediation." *The Journal of Finance*, 32 (2): 371-387.

Panyagometh, Kampohl, and Gordon S. Roberts. 2002. “Private Information, Incentive Conflicts, and Determinants of Loan Syndications." York University Working Paper 20:577-593.

Preece, Dianna, and Donald Mullineaux. 1996. “Monitoring, Loan Renegotiability, and Firm Value: The Role of Lending Syndicates." *Journal of Banking and Finance*, 20:577-593.

Qian, Jun, and Philip E. Strahan. 2007. “How Law and Institutions Shape Financial Contracts: The Case of Bank Loans." *The Journal of Finance*, 62 (6): 2803-2834.

Simons, Katerina. 1993. “Why Do Banks Syndicate Loans?" *New England Economic Review*, pp. 45-52.

Simunic, Dan A. 1980. “The Pricing of Audit Services: Theory and Evidence." *Journal of Accounting Research*, 18 (1): 161-190.

Simunic, Dan A. 1984. “Auditing, Consulting, and Auditor Independence." *Journal of Accounting Research*, 22 (2): 679-702.

Strahan, Philip E. 1999. “Borrower Risk and the Price and Nonprice Terms of Bank Loans." Staff reports 90, Federal Reserve Bank of New York.

Sufi, Amir. 2007. “Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans." *The Journal of Finance*, 62 (2): 629-668.

APPENDIX A

Table A.1: Summary statistics

Variable	Mean	St. Dev.	Min	Max
All-In-Drawn Spread	4.600	0.865	2.140	7.279
Audit Fee	0.002	0.002	0.000	0.011
HHI	3936	1950	1119	10000
No of Segments	5.846	3.120	1.000	13.000
Accounts Receivable	0.156	0.113	0.000	0.620
Inventory	0.137	0.134	0.000	0.692
Deal Amount	19.248	1.331	12.571	23.901
Maturity Length	7.039	0.668	4.997	7.980
Number of Lenders	1.878	0.902	0.000	3.526
Secured/Unsecured	0.375	0.484	0	1
Secured Dummy	0.642	0.480	0	1
General Covenant Dummy	0.633	0.482	0	1
Financial Covenant Dummy	0.721	0.448	0	1
Perf. Pricing Dummy	0.550	0.498	0	1
Term Loan	0.140	0.347	0	1
Revolver/Line	0.667	0.471	0	1
Takeover	0.053	0.224	0	1
Debt Repay.	0.051	0.220	0	1
EBITDA	0.088	0.087	-0.906	0.341
Total Assets	7.372	1.673	2.030	10.845
Tobin's Average Q	1.812	1.032	0.634	9.490
Sales	1.228	0.798	0.019	4.421
Dividends	0.012	0.020	0.000	0.154
Leverage	0.274	0.170	0	1.036
Debt Due In One Year	0.101	0.177	0.000	1.000
Cum. Stock Returns	1.035	0.171	0.356	2.436
St. Dev. Stock Returns	0.108	0.065	0.004	0.724
Debt Rating	8.832	7.561	0	23
Debt Rating Dummy	0.607	0.489	0	1
N		2791		

Table A.2: First Stage - All Instruments

	(1)	(2)	(3)	(4)	(5)	(6)
	All-in-Drawn Spread		**No. of Lenders**		**Maturity**	
No of Segments	0.0001***	(0.000)	0.0001***	(0.000)	0.0001***	(0.000)
Accounts Receivable	0.0013***	(0.000)	0.0014***	(0.000)	0.0014***	(0.000)
Inventory	-0.0001	(0.000)	-0.0001	(0.000)	0	(0.000)
HHI	-0.0000**	(0.000)	-0.0000**	(0.000)	-0.0000**	(0.000)
Deal Amount	-0.0001	(0.000)	0	(0.000)	0	(0.000)
Maturity Length	-0.0001**	(0.000)				(0.000)
Number of Lenders	0.0001**	(0.000)				(0.000)
Secured/Unsecured	0.0002**	(0.000)	0.0001*	(0.000)	0.0001*	(0.000)
Secured Dummy	0	(0.000)	0	(0.000)	0	(0.000)
General Covenant Dummy	-0.0001	(0.000)	-0.0001	(0.000)	0	(0.000)
Financial Covenant Dummy	0.0001	(0.000)	0.0001	(0.000)	0	(0.000)
Perf. Pricing Dummy	0	(0.000)	0	(0.000)	0	(0.000)
Term Loan	0	(0.000)	-0.0001	(0.000)	-0.0001	(0.000)
Revolver/Line >= 1 Year	-0.0001	(0.000)	-0.0003***	(0.000)	-0.0003***	(0.000)
Takeover	0.0001	(0.000)	0.0001	(0.000)	0.0001	(0.000)
Debt Repay.	-0.0001	(0.000)	-0.0001	(0.000)	0	(0.000)
Constant	0.0082***	(0.001)	0.0069***	(0.001)	0.0069***	(0.001)
EBITDA	-0.0019***	(0.001)	-0.0020***	(0.001)	-0.0020***	(0.001)
Total Assets	-0.0007***	(0.000)	-0.0007***	(0.000)	-0.0007***	(0.000)
Tobin's Average Q	0.0002***	(0.000)	0.0002***	(0.000)	0.0002***	(0.000)
Sales	0.0001***	(0.000)	0.0001***	(0.000)	0.0001***	(0.000)
Dividends	0.0043**	-(0.002)	0.0042**	(0.002)	0.0042**	-(0.002)
Leverage	0.0003	(0.000)	0.0003	(0.000)	0	(0.000)
Debt Due In One Year	0.0005**	(0.000)	0.0005**	(0.000)	0.0005**	(0.000)
Cum. Stock Returns	0.0002	(0.000)	0.0002	(0.000)	0	(0.000)
St. Dev. Stock Returns	0.0025***	(0.001)	0.0024***	(0.001)	0.0024***	(0.001)
Debt Rating	0.0001***	(0.000)	0.0001***	(0.000)	0.0001***	(0.000)
Debt Rating Dummy	-0.0007***	(0.000)	-0.0007***	(0.000)	-0.0007***	(0.000)
R-squared	0.475		0.474		0.474	
N	2971		2971		2971	

Odd numbered columns report coefficient estimates, even numbered columns report p-value. * p < .10, ** p < .05, *** p < .01

The Thin Blue Line:
Police Militarization and Violent Crime[1]

J. Britton Haynes Jr.[2] Alexander F. McQuoid[3]

ABSTRACT

Recent events such as the riots in Ferguson, Missouri, have highlighted the increased militarization of police forces in the United States. This paper utilizes a new dataset that covers all military equipment transfers between the Defense Logistics Agency and local police forces from 1990 to 2014 to consider the effect of increased militarization on crime. These transactions were conducted under the Department of Defense's 1033 Program and constitute a major transfer of capital resources to local police departments with nearly two billion dollars transferred in the form of surplus military equipment. To deal with concerns of identification, we instrument for participation in the 1033 program using state-level exposure to the military through federal military spending set by Congress. We find that increased capital transfers to states embodied in military equipment reduces total violent crime and violent crime subcategories. The effect is large for overtly militaristic equipment such as assault rifles, but also for less militaristic transfers such as communication equipment, implying that both enhanced capabilities as well as power projection are important drivers of violent crime reduction. In addition, we find no evidence of a labor input response through additional hiring of sworn police officers, indicating that the program resulted in a more capital-intensive police force. Further, we find that increased police militarization results in lower incarceration rates even after controlling for reduced crime rates, suggesting a broader law and order impact beyond just enhanced capabilities. The results make clear that increased police militarization in the United States has played a meaningful role in the reduction in violent crime observed over the last twenty-five years.

JEL classification: H56; H7; H72; H76;

---

[1] The views expressed in this paper are those of the authors and should not be attributed to the United States Naval Academy or Department of Defense. We thank Amit Khandelwal, John Huston, Jonathan Dingel, David Grad, Joao Salles, Paul Landefeld as well as seminar participants at Florida International University, Eastern Economic Association Annual Conference, Western Economic Association Annual Conference, and the Congressional Budget Office for feedback and comments. All errors are our own.

[2] Ensign, United States Navy (brittonhaynes2@gmail.com)

[3] Economics Department, United States Naval Academy, 589 McNair Road, Annapolis, MD 21402 (mcquoid@usna.edu)

1 INTRODUCTION

On the night of November 24th, 2014, men patrolled the streets armed with M-4s adorned in body armor and camouflage fatigues, while shielding their faces with gas masks. This scene seemed reminiscent of a patrol in Iraq, Afghanistan, or some other foreign war-zone. However, these heavily armed men were not soldiers, but police officers in Ferguson, Missouri. The police presence in Ferguson put a spotlight on the growing trend of police forces across the United States employing military grade equipment.

Much, if not all, of this equipment is loaned to police departments across the country through the Department of Defense (DoD) Excess Property Program 1033. 1033 has transferred nearly $2 billion worth of military equipment to law enforcement agencies since its roots were formed in the National Defense Authorization Act for Fiscal Years 1990 and 1991. This militarization of United States police departments has raised serious questions about the practical need and implications for police forces to have access to military equipment such as Mine Resistant Ambush Protected (MRAP) vehicles and military style uniforms. Supporters state that police may operate more effectively with access to military grade equipment and tactics, while detractors argue policing is fundamentally distinct from, and antithetical to, military operations. While many Americans found the militaristic police images in Ferguson startling and uncomfortable, only limited empirical work has been done to understand the implications of this massive militarization of local police forces.

In this paper, we utilize data for all 50 U.S. States since the introduction of the 1033 program in 1997 to study the impact on crime rates through 2013. The amount of military equipment transferred to local police departments varies wildly across states and over time. We exploit this information to identify the impact of increased militarization on violent and property crime rates, as well as additional aspects of law and order.

To preview our results, we find that after accounting for state heterogeneity, national secular crime trends, and selection into the 1033 program, increased police militarization does meaningfully reduce violent crime rates. In addition, we find no effect of the program on sworn police offer rates, which suggests the program works primarily through enhancing the capital-intensity of police officers. We find evidence for both a capabilities enhancement effect and a "projection of power" effect on reducing violent crime rates. Finally, we find that incarceration rates are lower in areas with more militaristic police forces, even after controlling for declines in crime rates, suggesting a broader impact on law and order. Based on these results, we conclude that increased police militarization in the United States has played a meaningful role in the reduction in violent crime observed over the last twenty-five years.

The primary identification issue is that military transfers are not randomly distributed across states, but may respond to current crime or expectations about future crime. For example, if those areas with the most crime or those with appropriate expectations about future crime are the localities requesting more militarization, the estimated impact of militarization will capture both the true effect of increased militarization as well as the selection bias. Since both the sign of the effect and the selection bias are unknown, this could lead to a variety of inferential mistakes about the true impact of police force militarization.

To address this concern, we deploy an instrumental variables approach based on state exposure to militarization through federal military spending. Federal military spending differs across states and over time, and is driven by historical and international military needs that are unlikely to be directly related to current state crime levels. Federal military spending is controlled by Congress and mostly set in advance, making it unresponsive to local crime.

However, the presence of federal military spending provides state exposure to military culture, language, personnel, and equipment. While federal military spending is not directly correlated with local crime, exposure through increased interaction with military customs and personnel influences awareness and interest in military equipment. Federal military spending is therefore a viable instrument for militarization of the local police force through the 1033 program.

Using a naive bivariate approach, we find no effect of increased police militarization on violent crime rates when only police militarization is included in the empirical specification. The estimated effect is essentially zero. This observation, however, conceals important features in the data. First, the inclusion of time trends is significant as the US as a whole has experienced a secular decline in crime rates. Second, states differ dramatically in their exposure to crime, and failing to account for these differences results in misleading inferences about the impact of increased militarization. When both state fixed effects and control variables identified in the previous literature as significant determinants of crime are included, the estimated effect of militarization is negative, but small. One might be tempted to conclude that the massive resources transferred via the 1033 program were thus wasted.

However, even after controlling for observables, there is a lingering concern that selection into the 1033 program has not been adequately accounted for in the analysis. In particular, it seems likely that high crime areas might be more interested and willing to seek out additional resources, including military equipment transfers through the 1033 program. This would create a positive correlation between militarization and crime rates, and result in a biased estimated coefficient compared to the true causal effect.

These concerns are born out when militarization is instrumented for using federal military spending. The IV specification results in estimates vastly more negative, suggesting that selection into the 1033 program is a significant factor that needed to be accounted for in the analysis. As a result, the true estimated effect of militarization is quite large and statistically significant for total violent crime as well as disaggregated violent crime categories such as murder, forcible rape, aggravated assault, and robbery. We find no robust, statistically significant effects on total property crime or property crime categories. Our results are robust to alternative specifications, sample periods, manpower instruments, and assumptions on error terms.

The plan for the rest of the paper is as follows. Section 2 provides a review of the relevant literature, while Section 3 provides a brief description of the history and evolution of the 1033 program. Section 4 introduces the data and descriptive statistics. Section 5 presents the empirical approach and results. Section 6 concludes.

2 Related Literature

Despite the controversial nature of the 1033 program, there has been only limited empirical evaluations of the program, or the more general impact of police militarization on crime. The historical evolution of police militarization has been documented in Balko (2006), Paul and Birzer (2008) and Hall and Coyne (2012), with a more extensive treatment found in Balko (2013). Starting with Becker (1968), economists have long considered the interactions between policing, punishment, and crime. While theory has helped to develop a better understanding of the possible channels through which policing and punishment might impact crime, the economic literature has predominantly focused on empirical analysis (see DiIulio (1996) and Cameron (1988)).

The best empirical work has shown that the 1033 program has had a significant and important impact on crime. The two foundational papers in the literature are Harris et al. (2017) and Bove and Gavrilova

(2017). Harris et al. (2017) look at the impact of the 1033 program on local interactions between police officers and citizens, and find that the program reduced the number of citizen complaints, assaults on police officers, and increased drug arrests. Bove and Gavrilova (2017) find that overall crime rates are reduced, although there is no effect on arrest rates. Further exploration of mechanisms leads them to conclude that military equipment works by improving the capabilities of law enforcement to deter crime. Our approach here is similar to Bove and Gavrilova (2017), although we consider a longer period of analysis and use a related, but distinct instrumental approach.

As we treat 1033 as a transfer of capital, parallels arise between it and other programs designed to provide additional resources to police forces, such as the Community Oriented Policing Services (COPS) program that was established with the Violent Crime Control and Law Enforcement Act of 1994. Two of the main components of the COPS program were the Universal Hiring Program (UHP) and the Making Officer Redeployment Effective (MORE) grants. COPS began at a similar time as 1033 (at the time, the program was known as Program 1208 as discussed in the next section) due to high levels of crime at the time. 1033 and COPS are similar in that they both work via transfers of resources to Law Enforcement Agencies (LEAs). While 1033 is a direct equipment transfer, COPS provides grants for hiring more officers through UHP and grants for purchasing equipment through MORE. Both resource transfer programs were designed in order to improve the functioning of police forces and ultimately reduce crime. COPS focused on manpower and common capital equipment, while 1033 introduces restricted military grade equipment.

The COPS program has been analyzed in depth in Evans and Owens (2007). Utilizing the COPS program as an instrument to test whether shocks in the number of police officers has any effect on crime rates, they find that the COPS program increased the size of police forces and that there is a negative relationship with COPS UHP grants and four crime categories: auto theft, robbery, burglary, and aggravated assault. The authors also found that MORE grants had a negative relationship with the previous four crime rates as well as larceny rates. This finding suggests that 1033 may also have a negative relationship with crime rates since both 1033 and the MORE grant portion of the COPS program were designed to enhance police capabilities via equipment.

There is a long tradition of trying to estimate the effect of more police on crime. Cornwell and Trumbull (1994) were an early attempt to use a panel data approach to control for unobserved heterogeneity when they estimated a model of crime using counties in North Carolina. However, even after accounting for time-invariant heterogeneity and common trends, dealing with selection bias has been the key challenge in this literature (see Marvell and Moody (1996)). Solving the selection bias inherent in understanding the effect of police resources on crime has proven challenging, spurring innovative solutions. Levitt (1997) uses election cycles to instrument for police hiring with the identifying assumption being that politicians up for an election have an incentive ”to be tough on crime” by hiring more police officers independent of the actual crime dynamics, although questions have been raised about the approach (see McCrary (2002), Levitt (2002), and Worrall and Kovandzic (2010)). Fisher and Nagin (1978) argues there are more fundamental endogeneity concerns when studying police force size and crime determination.

More recently, researchers have used features of terrorism (and institutional responses to terrorism) as instruments for police resources. Klick and Tabarrok (2005) use terror alert levels to instrument for police levels in DC to tease out the effect of police on crime. They find that higher levels of alerts, during which DC police are more highly mobilized, are associated with lower crime levels. Di Tella and Schargrodsky (2004) utilize the allocation of police forces after a terrorist attack in Buenos Aires to study the effect on motor vehicle theft under the assumption that police force allocation in response to a terrorist attack is unrelated to crime dynamics. They find that the deterrence effect of police is large, but highly local. Draca et al. (2011) follow a similar approach in the aftermath of the 2005 terrorist attacks in London, and find similar results.

There is further debate about determinants of crime beyond police resources. Levitt (2004) and Shoesmith (2010) provide an overview of the literature on the determinants of crime, and in particular, the decline starting in the 1990s. There is some consensus on appropriate demographic and employment factors. Ethnic and racial disparities in crime and the criminal justice system are discussed in Sampson and Lauritsen (1997), Howson and Jarrell (1987), and Resignato (2000). Macroeconomic cycles have been identified as important by Raphael and Winter-Ebmer (2001) while measures of income have been studied by Doyle et al. (1999). The role of the age distribution and its impact on violent crime has been documented in Perkins (1997), while Levitt (1998) focused on juvenile crime and punishment. Howsen and Jarrell (1987) discusses determinants of property crime, and shows that violent and property crime are influenced by different factors.

Prison populations are thought to have a negative relationship with crime as suggested by Marvell and Moody Jr (1994) among others. Incarceration is thought to work through two specific channels, an incapacitation effect (”criminals off the street”) and a deterrence effect through the increased threat of punishment. Levitt (1996) uses prison overcrowding and institutional features as a natural experiment to identify the causal effects of prison population growth on crime. Kuziemko and Levitt (2004) study the impact of imprisoning drug offenders specifically on the determination of crime.

Our identification strategy is reminiscent of Nakamura and Steinsson (2014) (see also Hooker and Knetter (1997)), who use the differential impact of military procurement at the state level to identify fiscal multipliers, and Barro and Redlick (2011) who use military spending changes related to wars to identify the effects of government spending and taxes on output fluctuations. In a cross-country context, Creasey et al. (2012) use variation in military foreign aid to study national building and growth. The political economy of military spending is highly political, as documented by Mintz (2002). This is particularly useful for our identification strategy for two reasons. First, national military spending is driven by geopolitical events rather than local crime dynamics. Second, there is significant variation in state exposure to military spending, and this is related to historical circumstances that are unlikely to respond to local crime changes (see Braddon (1995)). These two observations suggest that the exclusion restriction is valid, and as we show in the analysis below, federal military spending is highly correlated with intensity of participation in the 1033 program, making this a valid instrument for police force militarization.

Having surveyed the related literature, we turn next to the specific characteristics of the 1033 program.

3 Background: 1033 Program

With the National Defense Authorization Act for Fiscal Years 1990 and 1991, Congress set the foundation for what would later become the 1033 program. This act allowed the DoD to transfer surplus equipment to state and federal agencies to help fight the *War on Drugs* through the 1208 Program. The 1208 Program was operated directly from the Pentagon through the Regional Law Enforcement Support Offices. These offices facilitated the transfer of surplus military equipment from the DoD to state and federal law enforcement. In October of 1995 the Defense Logistics Agency (DLA), the primary supply agency of the Department of Defense, took control of the program. The National Defense Authorization Act for Fiscal Year 1997 expanded 1208 and re-branded it as the 1033 Program.

The transition from 1208 to 1033 greatly increased the size of the DoD transfer program. This expansion allowed for all law enforcement agencies to acquire property for bona fide law enforcement purposes that assist in their arrest and apprehension mission. Under 1033, the requirement for equipment to be used to support the War on Drugs was dropped, yet requests for materials to support counter drug operations still received priority. Following the transition to 1033, between 1997 and 1999 a National Program Office was created at DLA in order to oversee the entire program. In 2009 oversight of the program was transitioned to the newly founded DLA Disposition Services Law Enforcement Support Office (LESO) headquartered at Battle Creek, Michigan. This transition created an office dedicated to the execution and support of the 1033 program.

To become a beneficiary of the 1033 program a state must create a Memorandum of Agreement (MOA) with the DLA. Once an MOA is formed, the governor of the state must appoint a DLA State Coordinator who is charged with the oversight of the program within the state to include accountability and proper use of transferred equipment. Once a State Coordinator is assigned Law Enforcement Agencies (LEAs) may apply to participate in 1033. Currently, DLA has an MOA with all 50 states, D.C., and the territories of Puerto Rico, Guam, and the Virgin Islands. Once approved by the State Coordinator and DLA LESO, a representative from an LEA may visit a DLA Disposition Services Site or visit the DLA's online webpage to see what equipment is available for transfer. After visiting a DLA Site, a LEA may prepare a request for equipment accompanied by a justification for why the LEA needs said equipment. Equipment requests are then sent to the State Coordinator and then DLA LESO, who has the final say, for approval. If approved, a LEA may take charge of the property and is required to cover all costs associated with the transportation of the property.

Since the creation of the 1208-1033 program, nearly $2 billion worth of equipment has been transferred from the DoD to state and federal law enforcement. Figure 1 displays the yearly evolution of program transfers. These transfers take the form of a regulated loan of capital with the potential for later ownership. Once 1033 property has been transferred, LEAs are required to utilize the equipment within one year, and

must use the equipment for one year. After this initial year, LEAs are required to submit proof of possession in the form of pictures, descriptions, and serial numbers of all equipment transferred.

LEAs are only required to submit proof of possession annually for equipment valued over $20,000 or equipment that requires special demilitarization when it is no longer of use. Equipment that does not meet this criteria is assigned a DLA demilitarization (DEMIL) code of A. DEMIL Code A equipment does not need to be reported following the initial year of ownership, and LEAs are free to dispose of this equipment as they see fit after the initial year.[4] LEAs never receive de jure ownership of equipment valued over $20,000, or equipment that requires special demilitarization. LEAs that have been loaned these types of equipment must send them back to the DLA when they are deemed no longer useful by the LEA or by the State Coordinator. For items that are of low value and do not require demilitarization, LEAs take full ownership of the equipment after one year and are free to use and dispose of it as they see fit.

Accountability of property is a key feature of the 1033 program. Due to the nature of the items transferred, it is essential for states and the DLA to keep records of all DoD equipment in LEAs hands. To facilitate accountability and transparency, the DLA provides a roster of all equipment transferred from the DoD to LEAs from 1990 up to the present. The dataset is organized by state (or territory), and includes information on which specific agencies received items, what equipment was transferred, quantity received, value of transferred equipment (when purchased by DoD), date of transfer, and whether or not it requires special demilitarization or return to DLA.

For example, the Baltimore County Police Department received 275 M16A1 assault rifles (national stock number 1005-00-073-9421) with a total value of $137,225 on July 31st, 2007, with the requirement that they be returned to DLA for demilitarization. The rosters are not entirely military equipment though, as 1033 includes a large amount of Code A items. For example, the Washington County Police Department in Washington, GA, received three guitar amplifiers (national stock number 7720-00-415-1343) with a value of $583.50 each on February 26, 2014. This information forms the basis on the following analysis to study the impact of police militarization on crime.

Of the $2 billion of equipment transferred to LEAs through 1033, only 17.3% of it is equipment that has received a DEMIL code of A. Examples of these code A items include shredders, sleeping bags, cameras, tools, and various other construction or office items. The vast majority of the equipment, 82.7%, must be returned to the DLA for demilitarization. Examples of items with these codes include firearms, firearm components, firearm optics, tactical vehicles, aircraft, boats, and night vision or infrared equipment. One of the most expensive items in the code B-Q items is the “Mine Resistant Vehicle”. This nomenclature is assigned to a variety of vehicles commonly referred to by the military as Mine-Resistant Ambush Protected vehicles, or MRAPs. These vehicles were designed for the DoD to better protect service members in Iraq and Afghanistan from the threat of IEDs, small arms fire, and land-mines. Since the beginning of 1208-

[4] DEMIL codes are standardized and applied to all DLA equipment. Equipment that requires demilitarization is assigned a DEMIL code of B, C, D, E, F, G, or Q.

1033, 559 of these vehicles have been transferred to LEAs, ranging in value from $412,000 per unit to $1,309,299. Total MRAP transfers are valued at nearly $380 million, accounting for 22% of all code B-Q transfers and 18.3% of the entire program - more than all code A items combined.

Each locality participating in the 1033 program is in control of how much equipment they request from the DLA. Figure 2 displays the total value of transfers from 1990-2013 for each state via the 1033 program. Every state has received some form of transfer, although participation in the program varied widely. Florida is a clear outlier, with almost $300 million worth of transfers from 1990 to 2014. On the other end of the spectrum are the Virgin Islands (not depicted) and Alaska with $164,000 and $990,000 respectively. Summary statistics of total value of military transfers are provided in Table 1 below. Geographically, Southern states (including Washington D.C.) occupy seven of the top 10 states in terms of military transfers, with four in the top five: Florida, Texas, Tennessee, and Virginia.[5] Summary statistics for crime and military exposure are also reported in Table 1.

4 Data

The first step in the analysis is to define a measure of militarization of police forces. The source for this information comes from a DLA roster of all items transferred from the DoD to law enforcement agencies in all 50 states, DC, and three territories from 1990 to 2014. Our unit of militarization is the monetary value of military equipment provided by the DLA to each state per year in US Dollars. This metric is not representative of the total value of equipment provided to each state by the DLA, but only of equipment that receives a DEMIL code of B, C, D, E, F, G, or Q. This equipment was chosen as the metric of militarization due to the special demilitarization requirement that it be returned to the DLA.

Code A items are loaned to LEAs with the ability for them to have de jure ownership after one year, and most items are readily available on the civilian market and thus not truly “military”. Code A items were almost exclusively transferred from 2012-2014. The relatively short period for analyzing Code A items makes it inappropriate for inclusion in the primary analysis, although results are robust to the inclusion of these codes.

The remaining data is drawn from a variety of sources. The crime statistics for violent and property crime rates come from the FBI's Uniform Crime Reporting program (UCR). This data includes rates of crimes per 100,000 population for 50 states and Washington D.C. from the years 1990-2013. The data includes disaggregated measures of crime such as murder and non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny, and motor vehicle theft. The first four of these measures are aggregated up into a measure of Violent Crime (VCR), while the last three are grouped together as Property Crime (PCR). Of particular note is the great difference between the mean violent crime rate and property crime rate, with property crime being nearly eight times as prevalent as violent crime. Both rates

[5] Results are robust to the exclusion of these states, and therefore not driven by the states receiving the most military transfers.

vary widely across localities, which influences the decision to use a fixed effects model in order to account for locality specific heterogeneity. Furthermore, during the time period under consideration, crime rates have followed a secular decline nationally (see Figure 3), which motivates the inclusion of time trends in the analysis so as not to wrongly attribute secular forces to militarization.

Demographic data including total population, percentage of the population between 18-24, and percentage that is black comes from the U.S. Census. This data covers all 50 states and D.C. from 1990-2013 for population, and all states and D.C. from 2003-2013 for young and black percentages. Unemployment rates and median income data come from the Federal Reserve Economic Data (FRED) and covers all 50 states and D.C. from 1990-2013. The portion of a state's GDP devoted to the federal military comes from the Bureau of Economic Analysis (BEA) and covers all 50 states and D.C. from 1997-2013. Prison population statistics come from the Bureau of Justice Statistics for the years 1990-2013. Manpower data comes from Department of Defense personnel historical publications for 1997-2009, and is used for alternative instrument robustness confirmation.

Due to the limited or lack of available data for Guam, Puerto Rico, and Virgin Islands, these localities are not included in the analysis, while multiple sample periods are used due to limited availability of control variables. Where possible, we show that sample period selection has no qualitative effects on the results. Our primary model focuses on 1997-2013 and 2003-2013.

5 Empirical Analysis

The 1033 program allows for a unique analysis of the interaction between police and crime as it is focused on the capital of policing as opposed to the more common focus on the labor of policing. In what follows, we will first focus on model selection using the Violent Crime Rate, before turning to the Property Crime Rate and the seven disaggregated measures of crime rates. Along the way, we will highlight key aspects of the data that help to causally identify the effect of increased police militarization on crime.

We begin the analysis by conducting a simple bivariate regression between violent crime rates and our measure of police militarization, the value of restricted military equipment transferred to the state in a given year. Limitations in data force us to consider three different samples. First, we focus on all years between 1990 and 2013, which includes periods prior to the implementation of the 1033 program when procurement took place under the smaller and more drug-specific 1208 program. Compared to later years, there were far fewer transfers in terms of quantity and value. Our second sample focuses on the 1033 program only, from 1997 until 2013. Finally, because of limitations with available controls, we also study a more recent period between 2003 and 2013 when all relevant controls are available. One result of the paper is that sample selection does not change the fundamental results, although the magnitudes of the point estimates do differ.

Table 2 reports estimates for the three different samples. In column (1), we estimate a naive pooled bivariate regression without exploiting the panel structure of the data. For all three samples, the point estimate is positive, although the all year sample is not statistically significant. If we took this estimate

seriously, one would conclude that the increase in police militarization lead to an increase in the violent crime rate in the US during the 1033 program. For every 1 million dollars of military equipment transferred to local police departments in the US, violent crime increased by 1.3 to 1.8 crimes per 100,000 residents.

One concern is that at the same time that police militarization was on the rise, there were national secular downward trends in all crimes, both violent and property, that were unrelated to militarization. Figure 3 shows both violent and property crime rates decreased dramatically from 1990 through 2013, and follow very similar national trends. Crime rates in 2013 were nearly half of what they were in 1990. This national decline in crime is a first order feature of the data, and important to account for if a true causal impact of militarization is to be estimated. If the true effect of police militarization on crime is positive, national trends towards lower crime might bias the estimates downward. In column (2) of Table 2 yearly time trends are included. The estimated coefficients on police militarization increase, and all samples imply that for each additional million dollars spent on military equipment for local police forces, violent crime increased by 3 offenses per 100,000 residents.

An additional concern, however, is that police militarization is not randomly assigned to police departments, but rather requested based on characteristics of the police departments. That is, there is likely a selection bias in the transfer of police departments. While the direction of the selection bias could go in either direction, one plausible view is that areas with high crime or expected high crime are likely to seek out more resources, including both labor and capital. This would create a positive correlation between militarization and (expected) violent crime, which would bias the estimate and understate the true effect. One simple approach to attempt to account for selection bias is to pull out state-specific factors that do not vary over time using state fixed effects. This is done in Column (3) and Column (4), where the latter includes both year and state fixed effects.

The estimated coefficient changes dramatically once differences across states are accounted for. The implied effect is now negative rather than positive, suggesting that increased police militarization actually reduces violent crime rather than exacerbates it. The switch in sign of the estimated effect is consistent with a selection bias story whereby the areas with higher (expected) crime are also the areas which utilize more resources including capital (military) equipment. Column (4) adds time fixed effects to the state fixed effects, pushing the point estimates closer to 0, once again confirming that national trends in crime need to be accounted for as well as state differences. The point estimate in Column (4) is highly significant at the 1% and 0.1% level depending on the sample period.

Table 2 encapsulates the challenges facing an empirical analysis of police militarization and crime rates. Local police departments face very different situations across the country and over time, and participation in the 1033 program is an active choice. To correctly identify the true causal effect running from police militarization to crime rates requires a plausible strategy to eliminate the selection bias inherent in the problem. Our next attempt to deal with the selection bias is through the use of control variables that have been previously identified in the literature as relevant for understanding the dynamics of crime in the US.

One limitation of the results thus far is the possibility of state and time-varying omitted variables. Without the inclusion of control variables, it is impossible to determine the robustness and validity of the previous estimates. The problem is that identifying correct control variables is not without debate. In an attempt to establish robust estimates, control variables are added to the model accounting for economic indicators as well as demographic indicators. These results are detailed in columns (1) through (6) of Table 3.

To get a sense of the possible omitted variables problem, control variables are included one by one in Columns (1) through (5), while Column (6) includes all control variables. In Column (1), the state unemployment rate is included, which is negative and statistically significant. The point estimate on police militarization is slightly smaller at -0.7 (compared to -0.9 without any controls) and is statistically significant at the 1% level. Column (2) includes the median income of the state with no resulting change in the original point estimate, or its statistical significance. Column (3) includes the prison population percentage, which has been identified as an important determinant of crime in previous literature. The point estimate on police militarization becomes more negative and is statistically significant at 0.1%. Column (4) includes the percentage of the population that is black, while Column (5) includes the percentage of the population that is between 18 and 24. In both cases the point estimate is similar to the original estimate and highly significant at the 0.1% level. Finally, in Column (6), we include all controls variables at once. The net result is that the estimated coefficient on police militarization is highly significant, and only slightly smaller than the estimate that only include time and state fixed effects. The results imply that for each one million dollars of military equipment transferred to local police departments, violent crime decreased by 0.72 offenses. Police militarization, time and state effects, and control variables explain about 50% of the observed variation in violent crime rates.

As discussed above, however, the addition of controls is ultimately an unsatisfactory exercise, as the underlying selection issue is not directly addressed. Instead, our preferred approach looks to address the selection issue head on through the use of instrumental variables. The key identifying assumption here is that federal spending on military is set by Congress, often years in advance, and that spending differs across states for historical reasons. That is, federal military spending in a state varies over time and across states, but for reasons that have nothing to do with local crime in a given year. Instead, we hypothesize that federal military spending influences access and use of the 1033 program through exposure to military customs, personnel, and equipment, but does not directly influence local crime rates. Our preferred specification is therefore:

$$Crime_{st} = \alpha + \beta_{militarization} * Militarization_{st} + \beta_z * Controls_{st} + State_s + Year_t + \epsilon_{st}$$

where the dependent variable is a crime rate variable, militarization is measured as the value of transferred military equipment to a state in a given year, and controls are drawn from previous literature. Given

concerns over selection bias, we instrument for militarization using federal military spending, and focus on $\beta_{militarization}$.

To understand the impact of the IV approach, we begin implementation of the IV using federal military spending in a model with no control variables. Due to data limitations, the addition of the controls lowers the number years of data that may be used. Starting first with no controls and the longer time series, Column (1) of Table 5 replicates the regression from Table 2. In Column (2), using the IV, the estimated coefficient on police militarization becomes much more negative and continues to be highly statistically significant. Whereas the simple OLS specification estimates the effect to be a reduction of -1.7 violent offenses per million dollars of military equipment, the IV estimates is nearly 15 times larger at -22.45. The IV estimates tells a very different story than the OLS estimate. The IV estimates suggests that there is a very large and meaningful reduction in violent crime as a result of the 1033 program and the resulting increased militarization of local police forces.

To evaluate the viability of the instrument, we need to consider the first stage effect of federal military spending in a state, and its impact on the 1033 program. The first stage (Table 4) finds a statistically significant relationship between federal military spending at the state level and participation in the 1033 program. Furthermore, the first stage F statistics is 15, suggesting we are unlikely to have a weak instruments problem. The first stage significance is capturing the fact that exposure to military personnel, language and customs promotes awareness of the 1033 Program and the equipment available. Furthermore, the changes in military spending at the state level are correlated with new investments in military equipment. The product cycle of military equipment means that old equipment is now surplus to requirements, and therefore available for procurement through the 1033 program.

Based on the assumed validity of the exclusion restriction and the strength of the first stage, this suggests the instrument is appropriate and useful, and that the resulting IV estimates strongly suggest that the 1033 program of increased police militarization has had an economically and statistically significant effect on reducing violent crime in the US.

Without properly accounting for national trends in crime, state heterogeneity, or selection bias associated with participation in the 1033 program, one would be likely to conclude that the 1033 program has a positive or negligible negative effect on violent crime. The truth of the matter is much different. Once properly considered, the 1033 program of transferring military equipment to local police departments has a meaningful impact on the reduction of violent crime.

To confirm that this finding is robust, we next consider the effect on violent crime with controls and instruments over the shorter sample. Columns (3) and (4) replicates results from Tables 2 and 3. Columns (5) and (6) report results when police militarization is instrumented for using federal military spending by state. While the magnitudes are smaller, the general result is confirmed. Failure to properly account for selection severely understates the true effect of militarization on violent crime. The impact of increased police militarization through transfers of military equipment lead to a 12.6 reduction in violent crime offenses

per 100,000 residents for each million dollars transferred. The effect is statistically significant at the 0.1% level, with the model explaining 85% of observable variation in violent crime across US states over time.

To check the robustness of the results to alternative instruments, we next consider multiple instruments based on the same underlying logic. We augment the federal military spending data with information on military personnel and total Department of Defense personnel. The idea is that exposure to DoD personnel, both civilian and military, has an effect on access to the 1033 program, but has no direct effect on local crime.

Table 6 reports results using personnel data from 2003 to 2009. Column (1) reports a simple multivariate regression over this shorter time horizon. The estimated effects are more negative, but also more variable, and no longer statistically significant, likely due to the lack of data. However, when we instrument using federal spending in Column (2), the coefficient is nearly 15 times as large as in Column (1), and statistically significant at the 5% level. The difference between this estimate and that found in column (6) of Table 5 is driven in part by differences in sample size due to data limitation on personnel data, although given the standard errors we cannot statistically distinguish between the two estimates. Column (3) of Table 6 extends the analysis by including military personnel along with federal military spending, with no discernible effect on the estimation. Column (4) includes a third instrument of total DoD personnel (both military and civilian), resulting is a slightly smaller point estimate of -24.47 and with greater statistical significance.

Next, we consider the robustness of inference to alternative assumptions on the error term. Table 7 re-estimates the model of Column (6) in Table 5 under four different error term assumptions. Column (1) uses a simple homoskedastic assumption, and concludes that the effect is statistically distinct from zero at 5% significance. Column (2) uses a heteroskedasticity-consistent standard error estimator, resulting in smaller standard errors and significance at 0.1%. Column (3) clusters at the state level under the assumption of correlated shocks within a state, resulting is slightly larger standard errors when compared to the robust estimator, but still significant at 1%. Lastly, in Column (4), we allow for both spatial and temporal dependence using a Driscoll-Kray estimator. Geographic regions in the U.S. tend to be similar in their approach to 1033. The concern is that geographic interdependence is not properly accounted for using alternative error assumptions. The standard errors are smallest under this assumption, resulting in statistical significance at the 0.1% level. Regardless of the assumption on error terms, the inferential results are unchanged. Police militarization has a statistically significant reduction in violent crime.

The lessons from analyzing aggregate violent crime and police militarization are clear and robust. There is significant heterogeneity across states, and a national secular decline in violent crime, but increased police militarization via equipment transfers through the 1033 program are associated with significant (both statistically and economically) declines in violent crimes at the state level. While the estimated effect appears to be small if only control variables are included, the effect is noticeably larger once the selection issue is directly addressed using a valid IV approach. We find that increases in police militarization lead to lower violent crime rates.

Having established this result, we next consider the impact of militarization on property crime. There are good reasons to believe that while property crime and violent crime tend to move together over time, the specific effect of militarization could deter violent crimes while at the same time, push criminals into less confrontational crimes such as property crime. That is, militarization can alter the relative cost of committing a violent crime compared to property crime, not just the overall cost of committing crime. On the other hand, if a more militarized police force works primarily through a power projection effect, the impact on less aggressive property crime may be limited.

The results of Table 8 suggest that the impact on property crime is less robust then for violent crime. Columns (1) through (3) report simple bivariate regressions using the three different sample periods (as in Table 2 for violent crime) as well as state and year effects. In all three cases, the estimated coefficient is negative, although it decreases in magnitude as the sample size declines. In Column (3), which focuses on the period 2003-2013, the estimated effect is not significant. When controls are added in Column (4), the effect is reduced by 33\% and is statistically insignificant. As discussed above, controls do not directly address concerns over selection. Columns (5) and (6) re-estimate the models in columns (3) and (4) using an IV approach. The results are mixed. In both cases, the estimated coefficient is significantly more negative after accounting for possible selection in the 1033 program, as was the case with violent crime. However, statistical significance is far from robust. In column (5), with no additional covariate controls, the estimated impact of police militarization is significant at conventional levels. Adding in socio-economic controls in column (6) reduces the estimated coefficient, while increasing standard errors, resulting in a loss of statistical significance. While the point estimates are always negative, fragile inference prevents drawing strong conclusions about the impact of police militarization on aggregate property crime. However, the evidence does not support the view that police militarization is reduces certain types of crime at the expense of other types of crime.

While VCR and PCR are useful aggregates, the process of simple aggregation may be hiding important information about the determinants of crime, and how they respond to increased police militarization. To consider these potential differences, we separate VCR and PCR into their seven component crime rates to determine what specific offenses militarization may be providing deterrence for, or uncover any positive effects that may be hidden by focusing on the aggregate rate.

The results are displayed in Table 9. The violent crime rate is a simple aggregation of murder, forcible rape, robbery, and aggravated assault crime rates. Aggregate property crime includes burglary, larceny theft, and motor-vehicle theft. General patterns found at the aggregate level are confirmed when looking at the disaggregated components. First, there are no hidden reversals at lower levels of aggregation once selection is accounted for in the estimation. Second, the general estimation patterns uncovered for aggregate crime rates hold at disaggregated levels as well. When time and state fixed effects are included as well as controls, the estimated effects on police militarization are negligible even when they are statistically significant. However, using an instrumental variable approach results in noticeably larger and

more negative estimated effects. The consistency of the patterns at both aggregate and disaggregate levels is comforting.

All four violent crime categories are statistically significant, while all three property crime categories are insignificant. While the point estimates for all seven categories are negative, the disaggregated category results underscore that property crime effects are fragile while the police militarization impact on violent crimes is robust. By considering the disaggregated categories, we both confirm the robustness of the results (and the patterns of the results), and also discover the channels through which increased police militarization operates.

While the 1033 program is effective in reducing violent crime, to more fully evaluate the effectiveness of the program, we should consider whether it is the militarization that matters, or whether it is the additional resources that matter. One interpretation of the program is that it transferred significant capital resources to police departments in need of resources, and as such, one should expect a reduction in crime. An alternative interpretation is that specific military resources were transferred, and these military resources have specific effects on crime deterrence. Or put differently, was it the capital resources in general or the military resources in particular that affected crime.

We can try to provide a preliminary answer to this question by disaggregating transfers based on the National Stock Number (NSN). A NSN is a 13-digit code attached to each piece of equipment, broken up into a 4-digit Federal Supply Classification (FSC) followed by a 9-digit item code. The FSC groups together similar supply products. The first FSC of interest are codes in the 1000s, which include weapons and aircraft and account for about 40% of the value of equipment transferred. The second FSC of interest are codes in the 2000s, which includes marine and ground vehicles and accounts for another 40% of transfers. The third FSC of interest are codes in the 5000s which include tools, hardware, scaffolding and prefabricated structures, construction and building materials, and communication equipment. This third category accounts for 10% of transfers.

The three categories we focus on are important because they make up the vast bulk of transfers and are regularly transferred throughout the period under consideration, but also because they represent different degrees of militarization. The FSC codes in the 1000s reflect high militarization resource transfers in the sense that these are capital resources that have very specific military capabilities. The FSC codes in the 2000s include equipment that has high military capabilities (such as MRAPs), but also less military capabilities such as utility trucks and cars. The FSC codes in the 5000s include specialized and non-specialized equipment that provides capital resources without the heavy military emphasis. By disaggregating the data long measures of high and low militarization, we can try to estimate the relative impact of specific military equipment compared to general capital equipment transfers.

Results are reported in Table 10. In columns (1)-(3), the baseline regression with state and year fixed effects are reported with each of the three militarization variables included in the specification using the different sub-samples under consideration. Regardless of the sample period, medium military transfers have the largest (negative) impact on violent crime. Low military transfers have the next largest negative

effect, but the coefficient is not statistically significant. High military transfers have a consistently small negative and statistically significant effect on crime. The addition of controls in Column (4) doesn't alter the story noticeably. The impact on high and medium militarization is negligible, while the estimated coefficient on low military turns positive (although still statistically insignificant). These results suggest that it is the militarization of the equipment transfers that matters for violent crime reduction

The story changes, however, once selection concerns are addressed directly with instrumental variables. We focus on high and low militarization categories, which have the sharpest contrast in military capabilities. In columns (5)-(8), we instrument for the highlighted category only, both with and without control variables. Comparing column (5) to column (7), we see that the estimated impact for low military transfers is nearly 3 times as large as high military transfers, although both are large and significant. Comparing columns (6) and (8), which include additional controls, the point estimates are very similar and tell the same story. Low militarization equipment transfers have a noticeably larger effect on violent crime than high military transfers, although both are economically meaningful and statistically significant.

The result suggest that militarization in and of itself does matter for violent crime deterrence, but that local police departments would also benefit from additional capital transfers. In terms of a cost-benefit analysis, this suggests that a brand new dollar of spending on capital resources would provide more bang when transferred as low militarization equipment. However, given the nature of the 1033 program, which transfers depreciated surplus military equipment, such a comparison is not completely appropriate. The empirical evidence supports the view that surplus depreciated military capital equipment of all kinds does reduce crime rates in the US, making the 1033 program quite successful in achieving the stated mission of increased police effectiveness in deterring and reducing crime.

To further explore the channels through which the 1033 program impacts crime, we next consider the impact of the 1033 program on the other major input to police production besides capital: labor. It is possible that the additional resources transferred via the 1033 program are reducing crime because the program frees up additional resources that can be used to hire more sworn police officers, or alternatively the value of an additional police officer increases because of the increased capital stock. Although identification is challenging, the general consensus is that increased police officers do reduce the amount of crime in an area. It may be that the observed effects from the 1033 program are operating through a labor channel rather than a specific militarization channel.

To address this possibility, we explore the impact of capital transfers on the sworn police officer rate in the state (officers per 100,000 residents). In column (1) of Table 11 we find a positive correlation between the value of military transfers and police officer rate in a state. In column (2), we include year dummies to account for any national changes in police officer rates, increasing the estimated coefficient close to 1, which literally implies that for each additional million dollars in transferred equipment, the state hired 1 additional police officer per 100,000 citizens. In column (3), state fixed effects are included and the coefficient switches signs and loses statistical significance. This suggests that within states over time, states

that receive more military equipment through the 1033 program have fewer police officers per 100,000 citizens, but the effect is not statistically distinguishable from zero.

The inclusion of controls in column (4) confirms the lack of a relationship. To deal with the selection concerns highlighted above, we again instrument for the intensity of participation in the 1033 program using federal military spending, and while the estimated effect becomes larger in magnitude, it continues to be negative and statistically insignificant. Column (6) includes controls along with the IV, and while the estimated coefficient switches signs again, it is statistically insignificant. Taken together, these results suggest that increased capital equipment through the 1033 program had no effect on labor inputs.

The lack of a labor input response supports the view that militarization itself is important for the estimated reduction in crime. Police departments in the US are becoming more capital-intensive, and in particular, more military-capital-intensive as a result of the 1033 program, resulting in reduced crime. To confirm the lack of an effect from labor, column (7) includes the police officer rate as an additional determinant of violent crime, with no statistically distinguishable effect. The coefficient on police militarization is slightly more negative compared to the estimate when police officer rate is excluded (-15 vs. -16.3), although the two estimates are not statistically distinct.

Finally, we consider the interaction between police militarization and institutions of order via the incarceration rate. Incarceration rates are thought to have a negative effect on crime rates through two different channels. One is a deterrence effect through increased cost of punishment when caught, and the other is an incapacitation effect by taking criminals off the street.

In Table 12, we explore the relationship between increased police militarization and incarceration rates. In column (1), a simple bivariate relationship suggests that areas with higher police militarization also have higher incarceration rates, which is weakly significant at the 10% level. The inclusion of year effects in column (2) has a negligible effect, while the inclusion of state fixed effects resulted in a much smaller point estimate that is no longer significant, but still positive. In columns (4) and columns (5), the total crime rate and the lagged total crime rate are included to capture the relationship between crime and incarceration. The estimates imply that crime in the previous year is associated with higher incarceration rates in the current year, consistent with a lag in the judicial process. After controlling for the relationship between crime and incarceration in a state, there is no significant different effect for areas with greater police militarization.

Once we account for selection using our instrumental variables approach, we find that police militarization has a negative effect on incarceration rates. This effect is significant at the 5% level. Even after controlling for the relationship between crime rates and incarceration rates, police militarization results in lower incarceration rates. Taken together with the earlier results, this suggests that areas with increased police militarization have lower violent crime rates, and furthermore incarceration rates are lower even after accounting for the fact that crime rates are also lower. This suggests that police militarization is working by making police forces more productive, which lowers crime directly, but also changes the calculus of crime, resulting in an additional reduction in incarceration rates. Given recent concerns about the social costs and

externalities of incarceration, it is particularly notable that a more military capital-intensive police force reduces both violent crime and incarceration rates.

6 Conclusions

The 1033 program has come under greater scrutiny due to concerns over the impact of recent trends of police militarization. Until recently, most of the rhetoric was driven without regard to empirical evidence. In this paper, we find robust support for the view that increased police militarization does reduce violent crime rates significantly, consistent with the emerging literature.

Furthermore, we find that both high military equipment (assault rifles, MRAPs, aircraft) and low military equipment (communications equipment and specialized tools) reduce violent crime rates, with the estimated effect somewhat larger for capital transfers that enhance capabilities independent of any power projection and deterrence. In addition, increases in capital stock had no effect on sworn police offer rates, suggesting that the program is working through increased (military-) capital-intensive police forces. That is, police departments used the 1033 program to enhance the capital stock per officer rather than hire more officers to utilize the additional equipment. This more capable police force has reduced violent crime, and even after controlling for these reductions in crime, states with a more militarized police force also have lower incarceration rates, suggesting a broader impact on law and order institutions beyond police productivity.

Unlike previous results found in Bove and Gavrilova (2017), we do not find robust support for the view that police militarization reduces property crimes. Although our point estimates are similar, we do not have sufficient statistical support to conclude these effects are different from zero. Our approaches differ in a number of dimensions, including of level of aggregation, length of time series, and identification strategy. However, all of these differences are also true for our analysis of violent crime, for which we find much more robust support for the view that police militarization reduces these types of crimes. Caution is warranted in overstating the impact of police militarization on less violent types of crimes such as motor vehicle theft and larceny, and deserves further investigation.

A potential issue with the present approach concerns heterogeneity within states. There are two separate problems to consider. First, as discussed in Aneja et al. (2012), there are serious concerns about the reliability of crime data at sub-state levels. Maltz and Targonski (2002) argue that because of concerns over reporting, county-level crime statistics should not be used to evaluate the effects of policy changes. Maltz and Targonski (2003) argue that the state-level crime data are less problematic than county-level data due to the FBI's cleaning and imputation process.

The second issue concerns the spatial nature of crime and crime deterrence. Crime is highly local, often affecting neighboring street blocks differently. The process of aggregation to the state level may be missing important differences at lower levels. Pushing back on this concern is the fact that police equipment can be deployed easily across local jurisdiction. Much of the equipment transferred was to county and state

LEAs that operate across multiple local jurisdictions. Furthermore, inter-agency cooperation would generate additional spillovers of resources across LEAs.

The question ultimately becomes one of appropriate aggregation. Our view is that the state is the appropriate level of analysis because of the nature of the 1033 program, which runs through a state coordinator. Spillovers across jurisdictions within a state are likely to be large, while spillovers across states relatively small. Concerns over measurement error at sub-state levels further support the decision to focus on state outcomes. Compared to Bove and Gavrilova (2017), choice of aggregation does not seem to matter for violent crime, but may matter for property crime, possibly because property crime is more substitutable across space and more susceptible to measurement error.

While the stated objective of the 1033 Program is to provide capital resources to facilitate the mission of crime deterrence and reduction of law enforcement agencies, there may be additional effects of the program beyond crime reduction. In this paper, we focus solely on the effect on crime. However, additional dimensions of social well-being beyond crime may be impacted by increased police militarization. If increased police militarization increases social fracture, for example, the costs may exceed the benefits of reduced violent and property crime estimated here. Insler et al. (2018) find that police militarization reduces civic engagement (measured by charitable giving and volunteering) for black households, but has no impact on white households.

Nonetheless, the findings of this paper are important for understanding the impact increased police militarization has on crime rates. Without properly accounting for geographic heterogeneity, secular national declines in crime, and selection into the 1033 program, one would conclude that increase militarization had no effect or even increased crime. Once all of these factors are accounted for, however, it becomes clear that the 1033 has played a meaningful role in the reduction of violent crime in the United States.

References

Aneja, A., Donohue III, J. J., and Zhang, A. (2012). The impact of right to carry laws and the nrc report: The latest lessons for the empirical evaluation of law and policy. Technical report, National Bureau of Economic Research.

Balko, R. (2006). Overkill: the rise of paramilitary police raids in America.

Balko, R. (2013). Rise of the warrior cop: The militarization of America's police forces. Public Affairs.

Barro, R. J. and Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. The Quarterly Journal of Economics, 126:51–102.

Becker, G. S. (1968). Crime and punishment: An economic approach. The Journal of Political Economy, 76(2):169–217.

Bove, V. and Gavrilova, E. (2017). Police officer on the frontline or a soldier? The effect of police militarization on crime. American Economic Journal: Economic Policy, 9(3):1–18.

Braddon, D. (1995). The regional impact of defense expenditure. Handbook of defense economics, 1:491–521.

Cameron, S. (1988). The economics of crime deterrence: a survey of theory and evidence. Kyklos, 41(2):301–323.

Cornwell, C. and Trumbull, W. N. (1994). Estimating the economic model of crime with panel data. The Review of economics and Statistics, pages 360–366.

Creasey, E., Rahman, A. S., and Smith, K. A. (2012). Nation building and economic growth. The American Economic Review, 102(3):278–282.

Di Tella, R. and Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. The American Economic Review, 94(1):115–133.

DiIulio, J. J. (1996). Help wanted: Economists, crime and public policy. The Journal of Economic Perspectives, pages 3–24.

Doyle, J. M., Ahmed, E., and Horn, R. N. (1999). The effects of labor markets and income inequality on crime: evidence from panel data. Southern Economic Journal, pages 717–738.

Draca, M., Machin, S., and Witt, R. (2011). Panic on the streets of London: Police, crime, and the July 2005 terror attacks. The American Economic Review, pages 2157–2181.

Evans, W. N. and Owens, E. G. (2007). Cops and crime. Journal of Public Economics, 91(1):181–201.

Fisher, F. and Nagin, D. (1978). On the feasibility of identifying the crime function in a simultaneous equation model of crime and sanctions. Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates, pages 361–99.

Hall, A. R. and Coyne, C. J. (2012). The militarization of us domestic policing. The Independent Review: A Journal of Political Economy, Forthcoming, pages 12–50.

Harris, M. C., Park, J. S., Bruce, D. J., and Murray, M. N. (2017). Peacekeeping force: Effects of providing tactical equipment to local law enforcement. American Economic Journal: Economic Policy.

Hooker, M. A. and Knetter, M. M. (1997). The effects of military spending on economic activity: Evidence from state procurement spending. Journal of Money, Credit, and Banking, pages 400–421.

Howsen, R. M. and Jarrell, S. B. (1987). Some determinants of property crime: Economic factors influence criminal behavior but cannot completely explain the syndrome. American Journal of Economics and Sociology, 46(4):445–457.

Insler, M., Murray, B., and McQuoid, A. (2018). From broken windows to broken bonds: Militarized police and social fragmentation. Technical report, USNA Working Paper Series.

Klick, J. and Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. Journal of Law and Economics, 48(1):267–279.

Kuziemko, I. and Levitt, S. D. (2004). An empirical analysis of imprisoning drug offenders. Journal of Public Economics, 88(9):2043–2066.

Levitt, S. D. (1996). The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. The Quarterly Journal of Economics, 101(2):319–351.

Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. The American Economic Review, pages 270–290.

Levitt, S. D. (1998). Juvenile crime and punishment. The Journal of Political Economy, 106(6):1156–1185.

Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. American Economic Review, pages 1244–1250.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. Journal of Economic perspectives, pages 163–190.

Maltz, M. D. and Targonski, J. (2002). A note on the use of county-level UCR data. Journal of Quantitative Criminology, 18(3):297–318.

Maltz, M. D. and Targonski, J. (2003). Measurement and other errors in county-level UCR data: A reply to Lott and Whitley. Journal of Quantitative Criminology, 19(2):199–206.

Marvell, T. B. and Moody, C. E. (1996). Specification problems, police levels, and crime rates. Criminology, 34(4):609–646.

Marvell, T. B. and Moody Jr, C. E. (1994). Prison population growth and crime reduction. Journal of Quantitative Criminology, 10(2):109–140.

McCrary, J. (2002). Using electoral cycles in police hiring to estimate the effect of police on crime: Comment. American Economic Review, pages 1236–1243.

Mintz, A. (2002). The political economy of military spending in the United States. Routledge.

Nakamura, E. and Steinsson, J. (2014). Fiscal stimulus in a monetary union: Evidence from US regions. The American Economic Review, 104(3):753–792.

Paul, J. and Birzer, M. (2008). The militarization of the American police force: A critical assessment. Critical Issues in Justice and Politics, 1(1):15–29.

Perkins, C. A. (1997). Age patterns of victims of serious violent crime. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Raphael, S. and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. Journal of Law and Economics, 44(1):259–283.

Resignato, A. J. (2000). Violent crime: a function of drug use or drug enforcement? Applied Economics, 32(6):681–688.

Sampson, R. J. and Lauritsen, J. L. (1997). Racial and ethnic disparities in crime and criminal justice in the United States. Crime and Justice, pages 311–374.

Shoesmith, G. L. (2010). Four factors that explain both the rise and fall of us crime, 1970–2003. Applied Economics, 42(23):2957–2973.

Worrall, J. L. and Kovandzic, T. V. (2010). Police levels and rates: An instrumental variables approach. Social Science Research, 39(3):506–516.

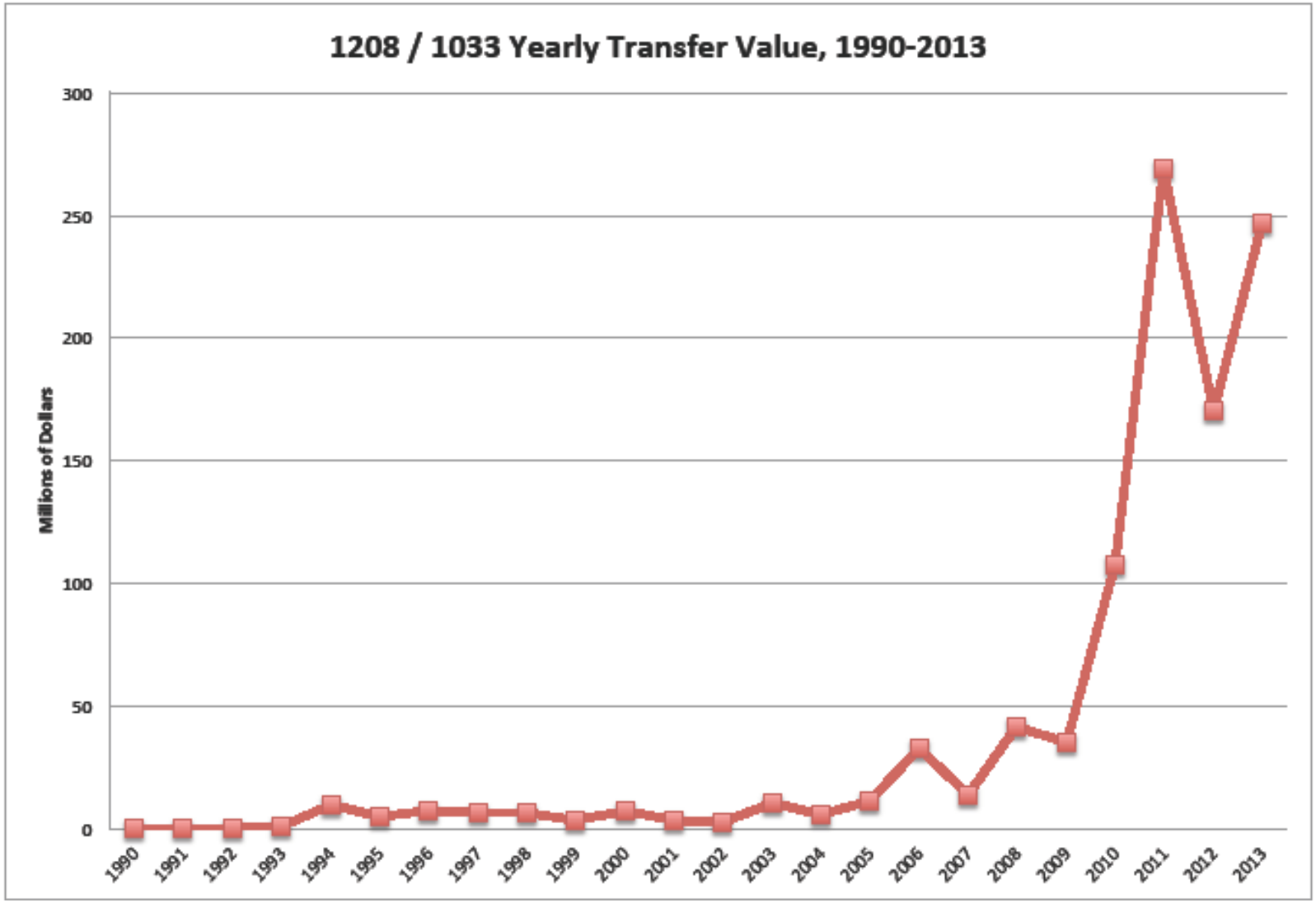


Figure 1: Program 1208 / 1033 Annual Transfer Values, 1990-2013

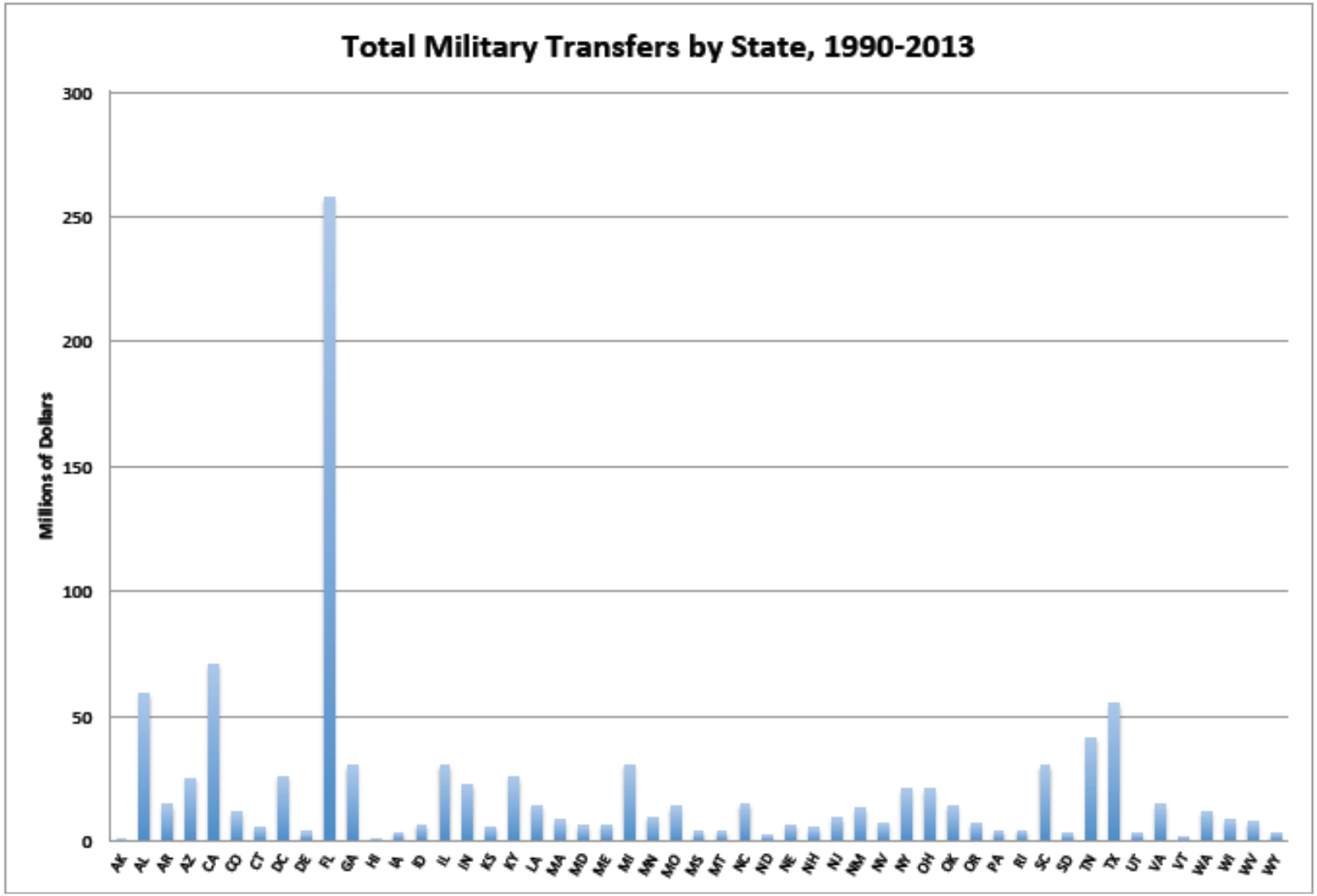


Figure 2: Total Transfers by State, 1990-2013

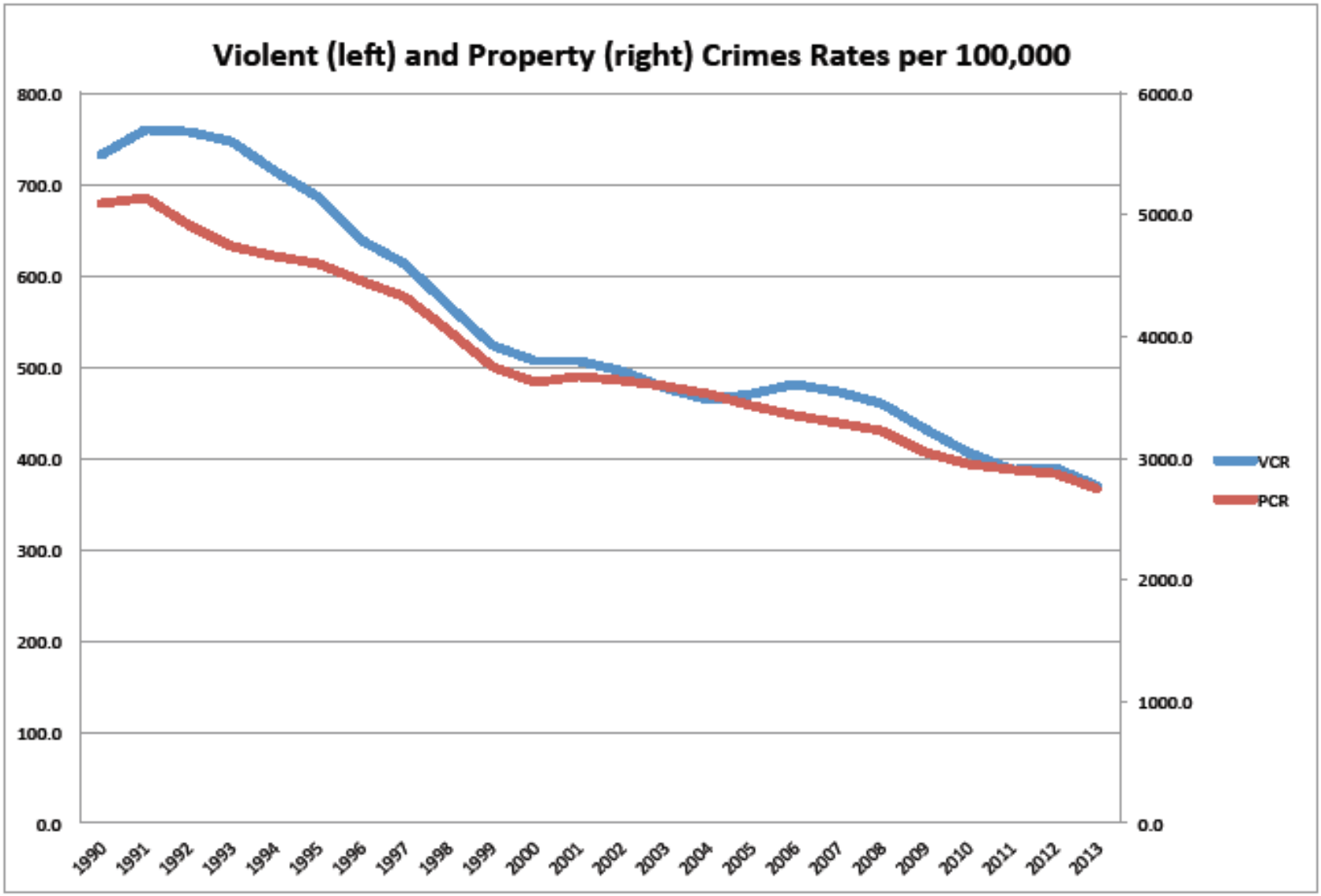


Figure 3: Violent and Property Crime Rates, 1990-2013

	Obs	Mean	Standard Deviation	Min	Max
Violent Crime Rate	1224	473.9	302.2	65.4	2921.8
Murder Rate	1224	6.2	7.1	0.2	80.6
Forcible Rape Rate	1224	35.4	12.8	9.7	98.6
Robbery Rate	1224	134.1	133.7	6.4	1266.4
Agg Assault Rate	1224	298.2	177.7	34.1	1557.6
Property Crime Rate	1224	3725.9	1153.5	1724.3	9512.1
Larceny Rate	1224	2554.1	731.1	1273	5833.8
Burglary Rate	1224	798.4	307.3	287.2	2170.6
Motor Vehicle Theft Rate	1224	373.4	242.7	53.3	1839.9
All Transfers	1224	0.85	5.24	0	154.79
Demil Code B-Q (restricted)	1224	0.80	5.14	0	154.79
FSC 1000 (Weapons and Aircraft)	1224	0.34	4.77	0	152.91
FSC 2000 (Marine and Group Vehicles)	1224	0.36	1.28	0	14.06
FSC 5000 (Tools, Construction Materials, Comm. Equip.)	1224	0.09	0.53	0	10.43
State GDP Federal Military (millions)	868	3337	4453	0	31391
Total DoD Personnel	816	33804.38	42972.6	577	243126
Military DoD Personnel	816	20646.58	29149.1	50	151945

Notes: Crime rates are measured as rates per 100,000. Violent Crime is an unweighted aggregate of four subcategories. Property Crime is an unweighted aggregate of three subcategories. All militarization variables are reported in millions of dollars. DoD personnel variables are number of personnel by state on September 30 of that year. Total DoD personnel is the sum of military and civilian DoD personnel.

Table 1: Summary Statistics: Crime and Militarization Variables

Violent Crime Rate	(1)	(2)	(3)	(4)	N
All Years 1990-2013	0.113 (0.706)	3.043 (1.585)*	-5.793 (2.515)**	-3.117 (1.010)***	1224
1033 Program 1997-2013	1.316 (0.710)*	2.924 (1.494)*	-3.180 (1.232)**	-1.729 (0.432)****	867
Sample Restriction: Available Controls 2003-2013	1.847 (0.917)**	2.839 (1.453)*	-1.856 (0.720)**	-0.899 (0.203)****	561
Year FE State FE	No No	Yes No	No Yes	Yes Yes	

Notes: Each cell reports the coefficient and standard error for police militarization in a regression with the violent crime rate as the outcome variable. Each row reports the estimated coefficient for a different sample selection period. Column 1 estimates a simple bivariate regression with no year or state fixed effects. Column 2 reports a simple bivariate regression with year but not state fixed effects. Column 3 estimates a simple bivariate regression with state but not year fixed effects. Column 4 estimates a simple bivariate regression with both state and fixed effects. State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent.

Table 2: Violent Crime and Militarization - Simple Bivariate Regression

Violent Crime Rate 2003 - 2013	(1)	(2)	(3)	(4)	(5)	(6)
Police Militarization	-0.702 (0.208)***	-0.877 (0.192)****	-1.008 (0.184)****	-0.869 (0.190)****	-0.893 (0.198)****	-0.795 (0.193)****
Unemployment Rate	-10.92 (4.444)**					-8.20 (4.130)*
Median Income		18.77 (14.81)				16.2 (10.43)
% Prison			353.2 (145.2)**			302.7 (143.8)**
% Black				17.77 (4.083)****		13.46 (8.927)
% Young					-3.999 (2.920)	-0.955 (2.136)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
N	561	561	550	561	561	550
within R^2	0.425	0.435	0.420	0.447	0.397	0.478

Notes: State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Prison population measurement not consistent in DC over sample period. As a result, DC is dropped from the analysis when Prison Population controls are utilized.

Table 3: Violent Crime and Militarization - with Control Variables

Police Militarization	(1) 1997-2013	(2) 1997-2013	(3) 1997-2013	(4) 2003-2013	(5) 2003-2013	(6) 2003-2009	(7) 2003-2009	(8) 2003-2009
Federal Military GDP (Millions)	0.000269 (0.000113)**	0.000228 (0.000101)**	0.000656 (0.000252)**	0.000904 (0.000293)***	0.000853 (0.000349)**	0.000341 (0.000165)**	0.000339 (0.000164)**	0.000443 (0.000165)***
Military (DoD) Personnel							-0.00000260 (0.0000107)	0.000125 (0.0000671)*
DoD Personnel								-0.000130 (0.0000639)**
Unemployment Rate					1.311 (1.174)	0.0749 (0.0425)*	0.0720 (0.0455)	0.0808 (0.0480)*
Median Income					0.314 (0.714)	-0.108 (0.125)	-0.115 (0.128)	-0.132 (0.125)
% Prison					25.16 (19.53)	2.791 (1.751)	2.732 (1.740)	2.489 (1.749)
% Black					-0.913 (0.664)	-0.00657 (0.0810)	-0.00322 (0.0797)	0.00615 (0.0853)
% Young					-0.0504 (0.170)	0.0716 (0.0420)*	0.0707 (0.0412)*	0.0725 (0.0447)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
N	867	867	867	561	550	350	350	350
R^2	0.04	0.10	0.21	0.26	0.28	0.63	0.63	0.64

Notes: State Clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. First stage results are reported for IV regressions presented in Tables 5 and 6. Three distinct sample periods are presented based on data limitations as discussed in paper. Police militarization is defined as value of transfers through 1033 Program in millions of dollars.

Table 4: Instrumental Variables: First Stage Results

Violent Crime Rate	(1) 1997-2013	(2) 1997-2013	(3) 2003-2013	(4) 2003-2013	(5) 2003-2013	(6) 2003-2013
Police Militarization	-1.729 (0.432)****	-22.45 (6.965)****	-0.899 (0.203)****	-0.795 (0.193)****	-15.61 (5.282)***	-15.02 (5.26)***
Unemployment Rate				-8.20 (4.13)*		12.35 (14.74)
Median Income				16.2 (10.43)		22.43 (12.98)*
% Prison				302.7 (143.8)**		537.24 (255.78)**
% Black				13.46 (8.93)		-1.87 (13.11)
% Young				-0.96 (2.14)		-0.465 (2.466)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrumented?	No	Yes	No	No	Yes	Yes
N	867	867	561	550	561	550
First Stage F	.	15.02	.	.	15.94	11.59

Notes: State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Prison population measurement not consistent in DC over sample period. As a result, DC is dropped from the analysis when Prison Population controls are utilized.

Table 5: Instrumental Variables: Federal Military State Spending (Violent Crime Rate)

Violent Crime Rate	(1) 2003-2009	(2) 2003-2009	(3) 2003-2009	(4) 2003-2009
Police Militarization	-5.53 (1.88)***	-28.31 (14.46)**	-28.67 (14.80)*	-24.88 (11.79)**
Unemployment Rate	-3.31 (3.79)	-0.92 (3.40)	-0.89 (3.41)	-1.28 (3.41)
Median Income	6.78 (8.91)	4.78 (8.72)	4.74 (8.74)	5.08 (8.59)
% Prison	243.1 (116.4)**	263.0 (117.5)**	263.3 (117.8)**	260.0 (115.9)**
% Black	14.1 (6.13)**	12.68 (5.43)**	12.7 (5.42)**	12.88 (5.44)**
% Young	-3.66 (2.56)	-1.11 (2.80)	-1.08 (2.812)	-1.50 (2.73)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Instrumented?	No	Yes	Yes	Yes
N	350	350	350	350
First Stage F	.	7.08	3.63	3.55

Notes: State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Prison population measurement not consistent in DC over sample period. As a result, DC is dropped from the analysis when Prison Population controls are utilized. Column (1) is standard regression. Column (2) instruments for police militarization using federal military component of state GDP. Column (3) instruments using first instrument plus number of DoD military personnel in state. Column (4) instruments using first two instruments plus total number of DoD personnel (military plus civilian).

Table 6: Alternative Instruments (Violent Crime Rate)

Violent Crime Rate	(1) Homoskedastic	(2) Robust SE	(3) Cluster SE	(4) Driscoll-Kray SE
Police Militarization	-15.02	-15.02	-15.02	-15.02
	(7.39)**	(4.381)****	(5.26)***	(1.15)****
Unemployment Rate	12.35	12.35	12.35	12.35
	(12.25)	(10.28)	(14.74)	(1.77)****
Median Income	22.43	22.43	22.43	22.43
	(21.42)	(10.88)**	(12.98)*	(6.97)***
% Prison	537.2	537.2	537.2	537.2
	(235.1)**	(177.4)***	(255.8)**	(76.95)****
% Black	-1.872	-1.872	-1.872	-1.872
	(14.77)	(9.05)	(13.11)	(4.10)
% Young	-0.465	-0.465	-0.465	-0.465
	(6.13)	(2.77)	(2.47)	(1.70)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Instrumented?	Yes	Yes	Yes	Yes
N	550	550	550	550

Notes: Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent.

Table 7: Alternative Standard Error Assumptions (Violent Crime Rate)

Property Crime Rate	(1) 1990-2013	(2) 1997-2013	(3) '03-'13	(4) '03-'13	(5) '03-'13	(6) '03-'13
Police Militarization	-7.898 (1.438)****	-3.374 (0.954)****	-1.275 (0.863)	-0.897 (1.00)	-54.14 (26.2)**	-45.80 (31.26)
Unemployment Rate				-30.23 (32.42)		34.69 (42.78)
Median Income				41.92 (71.38)		61.56 (56.02)
% Prison				1021.0 (968.7)		1761.7 (701.1)**
% Black				-8.62 (51.81)		-57.05 (43.38)
% Young				20.87 (18.28)		22.42 (15.37)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
IV	No	No	No	No	Yes	Yes
N	1224	867	561	550	561	550
First Stage F	.	.	.	.	15.94	13.47

Notes: State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Prison population measurement not consistent in DC over sample period. As a result, DC is dropped from the analysis when Prison Population controls are utilized.

Table 8: Property Crime Rate (PCR) and Police Militarization

	LHS	(1)	(2)	(3)	(4)
Violent Crime:					
	Murder	0.004	0.002	-0.006	-0.275
		(0.006)	(0.002)	(0.253)	(0.119)**
	Forcible Rape	-0.041	-0.021	-0.977	-0.836
		(0.016)**	(0.011)*	(0.390)**	(0.391)**
	Aggravated Assault	-0.651	-0.582	-9.990	-9.376
		(0.156)****	(0.128)****	(3.299)***	(3.336)***
	Robbery	-0.211	-0.194	-4.634	-4.523
		(0.0730)***	(0.0851)**	(1.886)**	(2.098)**
Property Crime:					
	Burglary	-0.332	-0.364	-9.053	-10.45
		(0.191)*	(0.234)	(6.103)	(7.669)
	Larceny	-0.510	-0.564	-36.32	-21.84
		(0.572)	(0.557)	(24.09)	(18.35)
	Motor Vehicle Theft	-0.433	0.0304	-8.767	-13.50
		(0.247)*	(0.286)	(12.07)	(8.703)
	Year FE	Yes	Yes	Yes	Yes
	State FE	Yes	Yes	Yes	Yes
	Controls	No	Yes	No	Yes
	IV	No	No	Yes	Yes
	N	561	550	561	550

Notes: Each cell reports the coefficient and standard error for police militarization in a regression where the outcome variable is listed in each row. Column 1 reports a regression with no control variables and no instruments. Column 2 reports a regression with controls (not reported), but no instrument. Column 3 reports a regression with no controls, but instruments for police militarization. Column 4 reports a regression with controls (not reported) and uses an instrument for police militarization. Control Variables are the same set used in all previous specifications. Year and State fixed effects are included in each regression. State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent.

Table 9: Disaggregated Violent and Property Crime Categories

Violent Crime Rate	(1) 1990-2013	(2) 1997-2013	(3) 2003-2013	(4) 2003-2013	(5) 1997-2013	(6) 1997-2013	(7) 1997-2013	(8) 1997-2013
High Military Weapons and Aircraft	-1.760 (0.303)****	-1.118 (0.107)****	-0.640 (0.0409)****	-0.520 (0.0815)****	-37.60 (15.83)**	-28.87 (10.13)***		
Medium Military Marine and Ground Vehicles	-13.92 (4.339)***	-6.890 (2.778)**	-4.612 (1.937)**	-4.42 (2.15)**				
Low Military Comm. Equipment, Tools, Building Materials	-10.57 (8.165)	-6.424 (6.404)	-2.066 (3.665)	0.820 (3.13)			-118.4 (26.87)****	-121.8 (33.95)****
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV	No	No	No	No	Yes	Yes	Yes	Yes
N	1224	867	561	550	867	867	867	867
R^2	0.682	0.672	0.659	0.690	0.928	0.929	0.928	0.932
First Stage F	.	.	.	.	4.68	8.55	27.08	16.11

Notes: State clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Regressions weighted by population. Column (4) includes all controls from earlier tables, while columns (6) and (8) only include controls for other militarization variables. Instrument for highlighted militarization variable only.

Table 10: Violent Crime and Militarization - High and Low Militarization

	(1) 2004-2013 Police Rate	(2) 2004-2013 Police Rate	(3) 2004-2013 Police Rate	(4) 2004-2013 Police Rate	(5) 2004-2013 Police Rate	(6) 2004-2013 Police Rate	(7) 2004-2013 VCR
Police Militarization	0.639	0.965	-0.173	-0.0971	-0.641	0.353	-16.33
	(0.242)**	(0.413)**	(0.118)	(0.130)	(1.758)	(2.788)	(6.804)**
Unemployment Rate				-3.867		-4.534	15.38
				(2.079)*		(5.085)	(19.69)
Median Income				1.974		1.768	26.03
				(7.237)		(7.686)	(13.00)**
% Prison				-5.460		-13.69	508.1
				(84.59)		(114.4)	(323.1)
% Black				12.82		13.37	-2.736
				(5.627)**		(6.830)*	(17.26)
% Young				1.568		1.565	-1.534
				(1.576)		(1.542)	(2.670)
Police Officer Rate							0.0363
							(0.129)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
State FE	No	No	Yes	Yes	Yes	Yes	Yes
IV	No	No	No	No	Yes	Yes	Yes
N	509	509	509	499	509	499	499
R^2	0.003	0.023	0.187	0.218	0.928	0.8380	0.965
First Stage F	.	.	.	.	9.76	9.11	8.54

Notes: Police Rate is number of police officers per 100,000 residents. State Clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Columns (1)-(6) focus on the determinants of the police officer rate, while column (7) considers the impact on Violent Crime Rate. Prison data is unavailable for District of Columbia, and therefore dropped from regressions in columns 4,6, and 7.

Table 11: Impact of 1033 Program on Sworn Police Officers

Incarceration Rate	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Police Militarization	1.996	1.948	0.209	0.290	0.304	-12.88	-13.43	-13.30
	(1.021)*	(1.062)*	(0.196)	(0.185)	(0.186)	(6.137)**	(5.727)**	(5.705)**
Total Crime Rate				0.0159	-0.00318		-0.00926	-0.0194
				(0.0102)	(0.00887)		(0.0203)	(0.0102)*
Lagged Total Crime Rate					0.0217			0.0117
					(0.0101)**			(0.0164)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
IV	No	No	No	No	No	Yes	Yes	Yes
N	850	850	850	850	850	850	850	850
R^2	0.006	0.016	0.965	0.966	0.966	0.968	0.968	0.969
First Stage F	.	.	.	.	.	12.60	8.68	10.18

Notes: Incarceration Rate is number of incarcerated per 100,000 residents. State Clustered standard errors in parentheses. Significance: * 10 percent; ** 5 percent; *** 1 percent, **** 0.1 percent. Total Crime Rate is the aggregate of the violent crime rate and property crime rate. The data covers 1997 until 2013 for all 50 states.

Table 12: Impact of 1033 Program on Incarceration Rates

State Dependence in the Natural Gas Price and Rig Count Relationship

Matthew Brigida*

ABSTRACT

The goal of this analysis is to understand the relationship between natural gas prices and the North American natural gas rig count. Prior research has generally not included changes in the rig count as a determinant of contemporaneous or future changes in natural gas prices. We first show that when not allowing state dependence, the rig count does not affect natural gas prices. We then show, when allowing state dependence in the natural gas and rig count relationship, changes in the rig count are an important determinant of future changes in gas prices when natural gas prices are high. That is, over our monthly sample from 1997 through 2013, we find that the rig count has a negative and significant relationship with future natural gas price changes (Granger-causes) when natural gas prices are above a $6.74/MMBtu threshold. However, when gas prices are below this threshold, then the rig count does not affect natural gas prices (though gas prices do affect the rig count). Moreover, we find evidence consistent with media reports that natural gas producers tend to 'kill any rally' in gas prices by markedly increasing gas production. These results are useful to any participant in the natural gas markets, particularly producers and utilities.

INTRODUCTION

This paper investigates the existence and extent of state dependence in the relationship between natural gas prices and the rig count, and the implications of such state dependence for explaining changes in these series. In particular, this analysis finds evidence that the rig count can be an important determinant of the changes in natural gas prices.

Prior literature on explaining the changes in natural gas prices has generally included the gas rig count only incidentally through an exogenous 'shutin' variable, which measures the proportion of natural gas production idled (mainly due to hurricanes in the Gulf of Mexico). Specifically, Ramberg and Parsons (2012), Brown and Yücel (2008), and Hartley, Medlock, and Rosthal (2008) include the 'shutin' variable as an exogenous determinant of the change in natural gas prices within the framework of an error-correction model. Notably, Ramberg and Parsons (2012) and Brown and Yucel (2008) sampled natural gas prices at the weekly frequency, which is too fine a partition for the change in the rig count to react to natural gas prices. The only exception is Brigida (2014) which included changes in the rig count as a control variable in an error-correction model of the relationship of natural gas and crude oil prices.

In this analysis, we first show that if you do not consider state dependence in the relationship between natural gas prices and the rig count, then you would come to the conclusion that natural gas prices affect the rig count, however the rig count does not affect natural gas prices. Thus the rig count would not need to be included as an explanatory variable for changes in natural gas prices. However, once you

consider state dependence (where the state is dependent on the lagged natural gas price) then the rig count significantly affects natural gas prices.

In related research, Boudoukh et al (2007) illustrated the importance of including state dependence in the structural relationship between an asset's returns and its fundamentals. They do so by showing temperature affects frozen concentrate orange juice futures returns only when the temperature is near freezing. If the state dependence is not accounted for, then it appears temperature has no effect on orange juice futures returns.

Their paper was in response to earlier work (particularly Roll (1984) and Roll (1988)) which found fundamentals explained generally little of an asset's returns. Roll (1984) and Roll (1988) did not include state dependence in the functional relationship between the asset and its fundamentals.

This analysis differs from Boudoukh et al (2007) in several ways. First, the threshold at which temperature affects orange juice is well known, whereas we must estimate the threshold in the natural gas price and rig count relationship. Second, temperature is exogenous, whereas natural gas prices and the rig count are both potentially endogenous variables necessitating the vector autoregression (*VAR*) form.

In recent research, Geng and Fan (2016) used a Markov regime-switching model to estimate the impact of the shale gas revolution on regional natural gas markets. This model assumes state-dependence. Similarly Potts and Yerger (2016) find structural breaks in the Pennsylvania oil and gas markets due to the increase in Marcellus shale production. Additional research has focused specifically on the changing state of the relationship of natural gas and crude oil prices (Ji et al (2018); Batten et al (2017); Zhang and Ji (2018); Zhu et al (2018)).

Lastly, understanding the determinants of changes in natural gas prices is of increasing importance because natural gas fired electricity generation is quickly increasing its share of the national power generation portfolio. These utilities buy their natural gas from natural gas production companies, but also often drill their own wells. Understanding the relationship between natural gas prices and the rig count will help these utilities to determine which types of electricity generation to use, whether to hedge natural gas prices, or drill their own wells. Further, many vehicle fleets (particularly city buses, etc.) have converted from diesel to natural gas. Therefore any change in natural gas prices will have an increasingly large effect on the macroeconomy.

MOTIVATION FOR STATE DEPENDENCE

We expect there to be state dependence between natural gas prices and the rig count driven by natural gas prices falling below marginal production costs. That is, when natural gas prices are near or below production costs, then the rig count will be highly dependent on natural gas prices as more costly rigs are idled. So the rig count is dependent on gas prices.

Alternatively, when natural gas prices are well above their production costs, increases in the rig count will cause a reduction in natural gas prices. That is, all rigs are brought online and thereby prices are tempered. In this case gas prices are dependent on the rig count.

A recent article in the financial press highlights the relationship between natural gas prices and the rig count. From the article:

Gas producers in North America including Chesapeake Energy Corp. (CHK) are killing their commodity's biggest rally in 10 months by opening more wells, putting the U.S. on track to have record gas supplies this year.

This is anecdotal evidence that above some threshold, natural gas producers bring enough rigs online to negatively affect gas prices.

The remainder of the paper is organized as follows. Sections 2 and 3 describe the data set and methodology respectively. Section 4 summarizes the results over various samples, and section 5 concludes.

DATA

Natural gas prices are logged monthly spot prices at the Henry Hub. Crude oil (west Texas intermediate) and heating oil prices are also logged monthly spot prices for delivery at Cushing, Oklahoma and the New York harbor (number 2) respectively. All price data are from the Energy Information Agency at the U.S. Department of Energy. All prices in this analysis are inflation-adjusted using the consumer price index for all urban consumers and all items.

North American rotary rig count data are from Baker Hughes. We use the Baker Hughes provided rig count split by natural gas and crude oil. The price and rig count series contain a unit root in levels, therefore first-differences are used.

Since the goal of this analysis is to investigate the relationship between natural gas prices and rig count, all data are sampled at the monthly frequency. Given costs associated with starting or idling a rig, it is highly unlikely drillers will react to natural gas prices at frequencies of less than a month.

Natural gas storage data are the monthly total working gas in storage series available from the EIA. The variable $STOR$ denotes the monthly deviation of the total working gas in storage from its 5-year average for that month. Notably the $STOR$ variable is a measure of actual storage amounts, which is different from the rig count. The change in the rig count reveals the expectations of market participants—specifically their expected future price relative to their privately-known production costs. Therefore, the rig count reveals new information to the market about these variables, which is not in other actual supply measures such as working gas in storage.

The heating and cooling degree-day (HDD and CDD respectively) data are population-weighted national averages and are available from the U.S. National Weather Service's Climate Prediction Center. In the later analyses we use $HDDdev$ and $CDDdev$, which denote the deviation of monthly total HDD and

CDD from their historical monthly norm. Using the augmented Dickey-Fuller test we are able to reject the null of a unit root at the 5% level for all data series we use in our later analysis (log differences in natural gas prices, crude oil prices, the rig count, $STOR$, $HDDdev$, and $CDDdev$).

Our full sample spans the months from February 1997 to June 2013. Given the effect of the 2008 financial crises on natural gas markets, we also use a January 2009 to July 2013 subsample.

METHODOLOGY

We will use a multivariate threshold vector autoregression (*VAR*) to model the relationship between logged natural gas prices and rig count. For more on threshold cointegration models see Balke and Fomboy (1997) and Tsay (1998).

Let $\boldsymbol{r}_t = (\Delta ng_t, \Delta rc_t)$ where $\Delta ng_t = ng_t - ng_{t-1}$ and $\Delta rc_t = rc_t - rc_{t-1}$. Then we have the multivariate threshold *VAR(p)*:

$$\boldsymbol{r}_t = \begin{cases} \boldsymbol{c}_1 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(1)} r_{t-1} + \sum_{i=1}^{p} \boldsymbol{\Delta}_i^{(1)} \boldsymbol{Z}_{t-1} + a_1^{(1)} & if\ ng_{t-1} \leq \gamma \\ \boldsymbol{c}_2 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(2)} r_{t-1} + \sum_{i=1}^{p} \boldsymbol{\Delta}_i^{(2)} \boldsymbol{Z}_{t-1} + a_1^{(2)} & if\ ng_{t-1} > \gamma \end{cases}$$

In this model the threshold variable is the prior period's logged natural gas price ng_{t-1}, and γ is the estimated threshold. $\boldsymbol{Z}_{t-1}$ is an optional vector of exogenous variables (described in section 4.4), and $a_t^{(i)}$ are independent sequences of two-dimensional white noise. The superscripts on the coefficient matrices refer to the estimated coefficients below (1) and above (2) the threshold. The estimation was done using the tsDyn (2009) package for the R programming language (2014). The threshold is estimated by finding the value, over the range of natural gas prices, which minimizes AIC. The value is found by a grid search. The estimation is done using conditional least squares.

Note that drillers certainly use expectations of future natural gas prices to decide whether to idle or bring a rig online, as opposed to using only present or past prices. The *VAR* incorporates these expectations so long as the drillers' expectations are formed using present and lagged values of the variables within the *VAR*. Hence, for robustness we also estimate the *VAR* with other variables (crude and heating oil, crude oil rig count, heating and cooling degree days and natural gas storage) which may affect expectations of natural gas prices.

RESULTS

In this section we present results over all our model specifications, and sample periods.

Full-Sample Without State Dependence

First we estimate a *VAR(2)* [4], without allowing for state dependence, on the full-sample of logged differences in natural gas prices and the rig count. The results are in Table 1 below. When testing for Granger-causation, we find that natural gas Granger-causes the rig count at the 0.1% level of significance, however the rig count does not Granger-cause natural gas prices. These results explain why earlier analyses of the determinants of the changes in natural gas prices have not included the rig count as an explanatory

variable.That is, without considering state dependence in the natural gas price and rig count relationship, you would conclude that the rig count does not affect gas prices.

Full-Sample Threshold *VAR(2)*

The results of the threshold *VAR(2)* estimated over the full sample period are shown in Table 2. First, the threshold was estimated to be $6.74/MMBtu (in July 2013 dollars). Below this threshold, both lagged changes in natural gas prices and the rig count positively and significantly affect present changes in the rig count. However, natural gas prices are unaffected by lagged changes in natural gas and the rig count.

Conversely, above the $6.74 threshold, lagged changes in the rig count have a negative and significant effect on natural gas prices. While above the threshold, natural gas prices have no effect on the rig count.

Table 1: Results of a VAR(2) on the full sample from February 1997 to June 2013*.*

	Δng_t	Δrc_t
constant	0.0014	-0.0008
	(0.0096)	(0.0023)
Δng_{t-1}	0.0521	0.0382
	(0.0726)	(0.0174)**
Δrc_{t-1}	-0.3424	0.6200
	(0.2932)	(0.0705)****
Δng_{t-2}	0.019334	0.0572
	(0.0706)	(0.0169)****
Δrc_{t-2}	0.329218	0.0651
	(0.2920)	(0.0702)
AIC	-1005.96	

*Table 1: Results of a VAR(2) on the full sample from February 1997 to June 2013. The sample size is 197 and there are 12 estimated parameters. denotes natural gas in $/MMBtu in July 2013 dollars, and denotes its natural log. denotes the natual log of the North American natural gas rig count. The standard errors are below the estimated coefficients in parentheses, and are heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West (1987)). *, **, ***, and **** denote statistical significance at the 10%, 5%, 1%, and 0.1% level respectively.*

So below the threshold, an increase in natural gas prices will tend to increase the rig count in subsequent months. Further the rig count has no effect on natural gas prices. However, above the threshold an increase in the rig count will tend to lower natural gas prices in the following months, and there is no effect of natural gas prices on the rig count. This is evidence that both the sign of the natural gas prices and rig count relationship, and the flow of causation, are state-dependent. Over our sample period, natural

gas prices were above the threshold for 33.83% of the months, and below for the remaining 66.16% of months.

Moreover, this evidence is consistent with the hypothesis that below the threshold there is a positive relationship whereby natural gas prices affect the rig count in subsequent periods. However, above the threshold the rig count is inversely related to subsequent changes in natural gas prices. This latter result is consistent with the anecdotal evidence that, in response to rallies in natural gas prices, producers increase the rig count to such an extent that natural gas prices are depressed.

The results above motivate tests for Granger-causation above and below the threshold. We can see in Table 3 that when natural gas prices are below the $6.74/MMBtu threshold, then changes in natural gas prices Granger-cause changes in the rig count. However rig count does not Granger-cause natural gas prices. Conversely, when natural gas prices are above its threshold, then rig count Granger-causes natural gas prices, but natural gas prices do not cause the rig count.

Table 2: Threshold VAR(2) Results on the full sample from February 1997 to June 2013.

Estimated Threshold: Natural Gas is $6.74/MMBtu				
	NG_{t-1} < $6.74		NG_{t-1} > $6.74	
	Δng_t	Δrc_t	Δng_t	Δrc_t
constant	0.0132	-0.0013	-0.0081	0.0035
	(0.0117)	(0.0029)	(0.0197)	(0.0049)
Δng_{t-1}	0.0620	0.0716	0.0799	-0.0002
	(0.0979)	(0.0241)**	(0.1033)	(0.0254)
Δrc_{t-1}	0.0850	0.6481	-2.8135	0.3478
	(0.3139)	(0.0773)***	(0.7730)***	(0.1904).
Δng_{t-2}	0.0975	0.0703	-0.1040	0.0314
	(0.0961)	(0.0237)**	(0.0989)	(0.0244)
Δrc_{t-2}	0.0108	0.0327	2.0608	0.2062
	(0.3105)	(0.0765)	(0.8028)*	(0.1977)
AIC	-2117.407			

The sample size is 197 and there are 21 estimated parameters. NG denotes natural gas prices in $/MMBtu in July 2013 dollars, and ng denotes its natural log. rc denotes the natual log of the North American natural gas rig count. The standard errors are below the estimated coefficients in parentheses, and are heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West (1987)). *, **, ***, and **** denote statistical significance at the 10%, 5%, 1%, and 0.1% level respectively.

Table 3: Tests for Granger-causality in the full-sample Threshold VAR(2).

Estimated Threshold: Natural Gas is $6.74/MMBtu		
	$NG < \$6.74$	$NG > \$6.74$
	F-Statistic	F-Statistic
$\Delta ng \Rightarrow \Delta rc$	13.4962	1.6963
	(0.0000)****	(0.1920)
$\Delta rc \Rightarrow \Delta ng$	0.2577	4.1683
	(0.7731)	(0.0201)**

⇒ denotes Granger-causation. ng denotes natural log of natural gas prices. rc denotes the natural log of the North American natural gas rig count. The p-value is below each estimated coefficient in parentheses. . **, ***, *** denoted statistical significance at the 10%, 5%, 1%, and 0.1% level respectively.

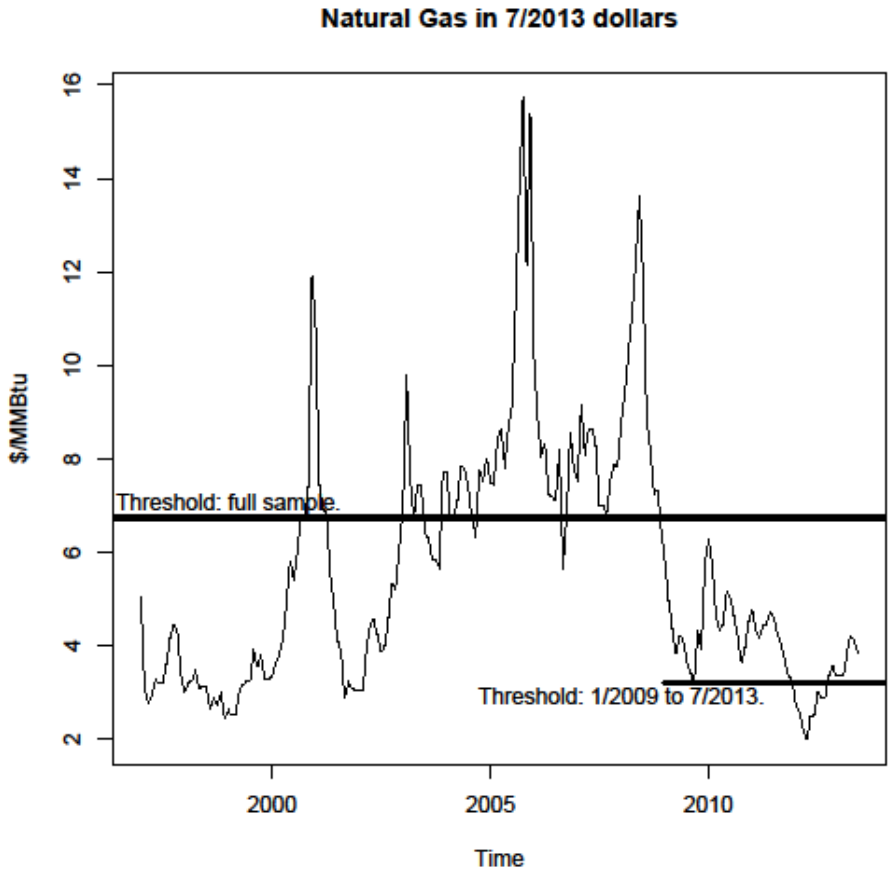
In sum, these results are evidence of both an endogenous relationship between natural gas prices and the rig count, and the existence of state-dependence in the nature of the relationship. They motivate the inclusion of the rig count, as a state-dependent variable, in analyses of natural gas price changes at the monthly sampling frequency and longer.

Subsample Threshold *VAR(2)*

During the 2008 financial crisis the (inflation adjusted) price of natural gas fell below $6/MMBtu, and has stayed below this value since. The 2008 crisis seems to have marked a structural shift in natural gas prices, and so we have estimated the threshold *VAR* over the period January 2009 to July 2013. Doing so, we find a lower threshold of $3.19/MMBtu. However, tests for Granger causation above and below the threshold are all insignificant. This is likely caused by too few data. Confirming the full-sample results on the data after the 2008 financial crisis will likely require more years of data.

The price of natural gas over the full sample, as well as the full-sample and subsample thresholds, are in Figure 1 below.

Figure 1: Natural Gas Prices with Thresholds



Full-Sample Threshold *VAR(2)* with Exogenous Variables

In this section we test whether our earlier results are robust to the inclusion of exogenous variables, which are commonly known to affect natural gas prices, in the *VAR*. The exogenous variables are the deviation of US working gas in storage from its 5-year average ($STOR$), the deviation of CDD and HDD from their long term norm ($CDDDev$ and $HDDDev$ respectively), and lagged changes in logged crude oil prices (west Texas intermediate).

Note, here we are considering crude oil prices as an exogenous variable, which has been a point of some debate. However, there is enough support (Serletis and Rangel-Ruiz (2004); Villar and Joutz (2006)) to herein treat crude oil as at least weakly exogenous.

The results are in Table 4 below. We find the earlier results of the threshold *VAR(2)* are largely robust to their inclusion. Below the price threshold, lagged changes in natural gas prices and the rig count positively and significantly affect changes in the rig count. However, these variables do not affect changes in natural gas prices.

Consistent with our main result, above the threshold lagged changes in the rig count negatively and significantly affect natural gas price changes. Also consistent is that natural gas prices do not significantly affect the rig count above the threshold.

The main difference from the earlier *VAR(2)* without exogenous variables, is that the natural gas price threshold is $8.05/MMBtu, as opposed to $6.74. Note, the maximum natural gas price over the period was $15.72/MMBtu.

Interestingly, below the threshold the natural gas storage variable inversely affects the change in natural gas prices (significant at the 1% level). However, above the threshold the storage variable is insignificant with respect to the change in natural gas prices, and the rig count variable is negative and significant.

This is evidence that, above and below the threshold, separate components of the natural gas supply chain affect prices. Below the threshold, the amount of gas in storage affects gas prices and not the rig count. Conversely, above the threshold rig count affects gas prices and not storage.

Full-Sample Threshold *VAR(2)* with Exogenous Variables and Production

Given the improvements in natural gas well technology in the last two decades, and the new shale plays this technology has enabled, there has been a marked increase in the amount of natural gas produced by *new* wells. We therefore estimate the above model, while controlling for the change in new well production per rig, and total production.

These data are gathered from the Energy Information Administration's monthly drilling productivity report, and are only available from January 2007 onward. The data are provided for key shale gas regions (Bakken, Niobrara, Anadarko, Permian, Eagle Ford, Haynesville, Appalachia), which we aggregate across regions.

We'll include the log differences of these variables in our model. Using the augmented Dickey-Fuller test, we are able to reject a unit root at the 5% level in both series. Regressing the log difference of the total change in production on the log difference of the production per new well, we estimate a slope coefficient of 0.3877 (significant at 1%), reflecting that increased new well production per well increases total production. Roughly, a 1% increase in new well production per rig increases total production by 0.38%. One important caveat is that this is only over those regions in the US that are shale gas plays. This is where the increase in technology will have the greatest effect, and excludes other gas supply sources where this relationship is likely to be less significant.

The estimated threshold of $3.3378 is lower than the $8.05 estimated in the previous model, which is largely due to lower overall natural gas prices in the later sample period. The AIC when including the production variables is -700.5387, compared with -1136.5710 when excluding the variables. The inferior AIC value of the model including the production variables is likely due to the smaller sample period, and more parameters estimated. The full results of the model are in table 5 below.

CONCLUSION

This analysis has shown the importance of considering state dependence in the determinants of natural gas prices. We first confirm that if you fail to consider state dependence, then you would erroneously conclude that the rig count has no effect on natural gas prices.

However, taking into account state dependence, as natural gas prices increase above a price threshold changes in the rig count negatively and significantly affect (Granger-cause) subsequent changes in natural gas prices. This means natural gas prices decline in response to an increase in the rig count. Notably, above the threshold changes in natural gas prices have no effect on subsequent changes in the rig count.

Conversely, below the price threshold, changes in natural gas prices positively and significantly affect (Granger-cause) subsequent changes in the rig count. However, changes in the rig count do not affect subsequent changes in natural gas prices.

This evidence is consistent with media reports stating natural gas producers often 'kill the commodity's rally'. That is, as natural gas prices rise while below the threshold, rigs with increasing marginal production costs are brought online in response. When prices rise above the threshold, and a large proportion of rigs are potentially profitable, enough supply is provided to negatively affect gas prices.

These results have practical implications for understanding changes in gas prices, and for implementing future models thereof. Further, these results are likely of interest to natural gas producers, who have much to lose if they are the last to bring a rig online in the face of declining gas prices. This analysis may help producers identify the price threshold above which the rig count will negatively affect gas prices, and thereby choose to not increase the rig count above this price. Ultimately such behavior may afford producers greater profits, and moderate volatility in natural gas prices.

Table 4: Threshold VAR Results on the full sample from February 1997 to June 2013 with the inclusion of exogenous variables.

Estimated Threshold: Natural Gas is $8.05/MMBtu				
	NG_{t-1} < $8.05		NG_{t-1} > $8.05	
	Δng_t	Δrc_t	Δng_t	Δrc_t
constant	0.0015	0.0025	-0.0152	0.0122
	(0.0128)	(0.0034)	(0.0407)	(0.0109)
Δng_{t-1}	-0.0122	0.0764	-0.3730	-0.0235
	(0.0886)	(0.0238)**	(0.2826)	(0.0760)
Δrc_{t-1}	-0.0423	0.5782	-5.7693	0.2356
	(0.2800)	(0.0753)***	(1.3409)***	(0.3609)
Δwti_{t-1}	-0.0611	0.0047	-0.1215	0.0845
	(0.1170)	(0.0315)	(0.3711)	(0.0999)
$STOR_{t-1}$	-0.0006	7.9e-05	2.0e-05	-3.0e-05
	(0.0001)***	(3.9e-05)*	(0.0005)	(0.0001)
$CDDdev_{t-1}$	-0.0007	0.0001	0.0043	-9.8e-05
	(0.0006)	(0.0001)	(0.0016)**	(0.0004)
$HDDdev_{t-1}$	-0.0006	8.8e-05	0.0016	-8.6e-05
	(0.0003)*	(7.4e-05)	(0.0007)*	(0.0002)
Δng_{t-2}	-0.0136	0.0553	0.4601	0.0221
	(0.0834)	(0.0224)*	(0.2474).	(0.0666)
Δrc_{t-2}	0.1535	0.0718	2.2238	-0.5860
	(0.2773)	(0.0746)	(1.9297)	(0.5193)
Δwti_{t-2}	0.2201	0.1224	0.9746	0.0402
	(0.1183)*	(0.0318)***	(0.3190)**	(0.0858)
$STOR_{t-2}$	0.0006	-9.1e-05	0.0001	1.7e-05
	(0.0001)***	(3.6e-05)*	(0.0005)	(0.0001)
$CDDdev_{t-2}$	-2.2e-05	-0.0001	-0.0046	-1.2e-05
	(0.0005)	(0.0001)	(0.0015)**	(0.0004)
$HDDdev_{t-2}$	-0.0003	-1.4e-05	-0.0010	6.4e-05
	(0.0002)	(6.4e-05)	(0.0006).	(0.0002)
AIC	-1136.5710			

The sample size is 192 and there are 53 estimated parameters. NG denotes natural gas prices in $/MMBtu in July 2013 dollars, and ng denotes its natural log. rc denotes the natual log of the North American natural gas rig count. The exogenous variables are: wti the natural log of crude oil prices; $STOR$ the deviation of US working gas in storage from its 5 year average for each month; $CDDdev$ and $HDDdev$ denote the deviation of monthly CDD and HDD from their long-term norm. The standard errors are below the estimated coefficients in parentheses, and are heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West (1987)). *, **, ***, and **** denote statistical significance at the 10%, 5%, 1%, and 0.1% level respectively.

Table 5

Estimated Threshold: Natural Gas is $3.34/MMBtu				
	NG_{t-1} < $3.34		NG_{t-1} > $3.34	
	Δng_t	Δrc_t	Δng_t	Δrc_t
constant	-0.1030	-0.0370	-0.0718	-0.0043
	(0.0654)	(0.0149)*	(0.0302)**	(0.0069)
Δng_{t-1}	0.2875	-0.0592	0.0465	0.1034
	(0.4982)	(0.1134)	(0.1644)	(0.0374)***
Δrc_{t-1}	1.1636	0.3772	-1.1484	0.9945
	(1.1734)	(0.2670)	(0.6085)**	(0.1385)****
Δwti_{t-1}	0.2086	-0.0105	-0.1685	0.0299
	(0.4788)	(0.1090)	(0.1864)	(0.0424)
$STOR_{t-1}$	-0.0002	5.5e-05	-0.0002	4.3e-05
	(0.0004)	(8.3e-05)	(0.0003)	(6.6e-05)
$CDDdev_{t-1}$	0.0008	0.0001	-0.0010	-3.3e-05
	(0.0013)	(0.0003)	(0.0008)	(0.0002)
$HDDdev_{t-1}$	0.0000	0.0002	-0.0013	-0.0001
	(0.0011)	(0.0002)	(0.0005)**	(0.0001)
$RigProd_{t-1}$	3.9429	-0.7594	0.1280	-0.4883
	(2.4917)	(0.5670)	(0.7847)	(0.1786)***
$TotProd_{t-1}$	-2.5432	0.5031	-0.1420	0.4822
	(2.4585)	(0.5594)	(1.1466)	(0.2609)*
Δng_{t-2}	-0.7174	0.1872	0.2303	-0.0125
	(0.6363)	(0.1448)	(0.1619)	(0.0368)
Δrc_{t-2}	3.8545	0.5203	0.9652	-0.2767
	(1.7305)*	(0.5594)	(0.6264)	(0.1426)*
Δwti_{t-2}	-0.0288	-0.0513	0.5266	0.0301
	(0.5267)	(0.1199)	(0.2047)**	(0.0466)
$STOR_{t-2}$	0.0006	-0.0001	0.0005	-2.2e-05
	(0.0003)*	(6.8e-05)	(0.0003)*	(6.5e-05)
$CDDdev_{t-2}$	-0.0020	0.0001	-8.3e-05	9.7e-06
	(0.0013)	(0.0003)	(0.0007)	(0.0002)
$HDDdev_{t-2}$	-0.0011	0.0002	-0.0001	0.0001
	(0.0012)	(0.0003)	(0.0004)	(9.2e-05)
$RigProd_{t-2}$	5.9565	1.3224	-0.9956	0.1426
	(4.2343)	(0.9635)	(0.8141)	(0.1852)
$TotProd_{t-2}$	1.1991	0.0190	0.9579	-0.2928
	(1.7251)	(0.3926)	(1.0594)	(0.2411)
AIC	-700.5387			

Threshold VAR Results on the full sample from February 1997 to June 2013 with the inclusion of exogenous variables. The sample size is 192 and there are 53 estimated parameters. NG denotes natural gas prices in $/MMBtu in July 2013 dollars, and ng denotes its natural log. rc denotes the natual log of the North American natural gas rig count. The exogenous variables are: wti the natural log of crude oil prices; $STOR$ the deviation of US working gas in storage from its 5 year average for each month; $CDDdev$ and $HDDdev$ denote the deviation of monthly CDD and HDD from their long-term norm. RigProd denotes the log difference in new production per rig, and TotProd denotes the log difference of total production. The standard errors are below the estimated coefficients in parentheses, and are heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West (1987)). *, **, ***, and **** denote statistical significance at the 10%, 5%, 1%, and 0.1% level respectively.

ENDNOTES

***** Associate Professor of Finance, SUNY Polytechnic Institute, 100 Seymour Rd, Utica NY 13502. email: matthew.brigida@sunyit.edu

[1] http://www.bloomberg.com/news/2012-11-14/gas-prices-doomed-to-stay-low-as-producers-pump-faster.html

[2] For robustness we also use prices for Brent north sea.

[3] http://ir.eia.gov/ngs/ngs.html

[4] The number of lags in the *VAR* was chosen using AIC

REFERENCES

Antonio, Fabio Di Narzo, Jose Luis Aznarte and Matthieu Stigler (2009). tsDyn: Time series analysis based on dynamical systems theory

Balke, N.S., and Fomby, T.B. (1997). Threshold cointegration. *International Economic Review* **38**: 627–645.

Batten, Jonathan A. and Ciner, Cetin and Lucey, Brian M. (2017) The Dynamic Linkages between Crude Oil and Natural Gas Markets. *Energy Economics* 62, 155–170. http://dx.doi.org/10.2139/ssrn.2653718

Boudoukh, J., M. Richardson, Y. Shen, R.F. Whitelaw (2007). Do asset prices reflect fundamentals? Freshly squeezed evidence from the OJ market. *Journal of Financial Economics* **83**: 397–412.

Brigida, M. (2014). The switching relationship between natural gas and crude oil prices. *Energy Economics* **43**: 48–55.

Brown, S.P.A. and M.K. Yücel (2008). What drives natural gas prices? *The Energy Journal* **29**: 45–60.

Geng, Jiang-Bo and Ji, Qiang and Fan, Ying (2016) The impact of the North American shale gas revolution on regional natural gas markets: Evidence from the regime-switching model *Energy Policy* 96, 167–178. http://dx.doi.org/10.1016/j.enpol.2016.05.047

Hartley, P.R., K.B. Medlock III, and J.E. Rosthal (2008). The relationship of natural gas to oil prices. *The Energy Journal* **29**: 47–65.

Ji, Qiang and Zhang, Hai-Ying and Geng, Jiang-Bo (2018). What drives natural gas prices in the United States?—A directed acyclic graph approach *Energy Economics* 69, 79-88. http://dx.doi.org/10.1016/j.eneco.2017.11.002

Newey, W. and K. West (1987). A simple positive semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**: 863–898.

Potts, Todd B. and Yerger, David B. (2016) Marcellus Shale and structural breaks in oil and gas markets: The case of Pennsylvania *Energy Economics* 57, 50–58. http://dx.doi.org/10.1016/j.eneco.2016.04.017

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ramberg, D.J. and J.E. Parsons (2012). The weak tie between natural gas and oil prices. *The Energy Journal* **33**: 13–35.

Roll, R. (1984). Orange Juice and weather. *American Economic Review* **74**: 861–880.

Roll, R. (1988). R^2. *Journal of Finance* **44**: 1-17.

Serletis, A., and R. Rangel-Ruiz (2004). "Testing for Common Features in North American Energy Markets," *Energy Economics* **26(3)**: 401–414.

Tsay, R.S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1202.

Zhang, Dayong and Ji, Qiang (2018) Further evidence on the debate of oil-gas price decoupling: A long memory approach. *Energy Policy* 113, 68–75. http://dx.doi.org/10.1016/j.enpol.2017.10.046

Zhu, Fangfei and Zhu, Yabei and Jin, Xuejun and Luo, Xingguo (2018) Do spillover effects between crude oil and natural gas markets disappear? Evidence from option markets *Finance Research Letters* 24, 25–33. http://dx.doi.org/10.1016/j.frl.2017.05.007

Could an Alternative Policy Design Have Produced a Stronger Mortgage Modification Outcome for HAMP?

Sean MacDonald, New York City College of Technology

ABSTRACT

This paper conducts a study of the relative effectiveness of the Home Affordable Modification Program (HAMP) - the primary federal mortgage loan modification program - from early 2009 through 2016. It evaluates U.S. Treasury Department and other data sources, and reviews the recent literature on the relative success of the program. The analysis suggests that HAMP's success rate in modifying mortgage loans was likely constrained by its voluntary design, a structure that enabled lenders and servicers to prioritize the interests of investors in assessing the risks of modification. It then considers the economic issues surrounding the foreclosure issue and presents a theoretical analysis, posing an alternative model illustrating where modification can be cost reducing. Concluding remarks reflect on the importance of promoting economic stability in policy design.

1. INTRODUCTION AND RESEARCH OVERVIEW

Despite indicators of emerging recovery in the U.S. housing market at the end of 2012, the problem of default and foreclosure remained a significant drag on economic recovery and job growth through 2012. This was particularly the case in many distressed housing markets nearly six years after the nation's foreclosure crisis began following the unraveling of the subprime mortgage market and the housing market collapse. Numerous policies to stem the rapid growth in foreclosures were introduced and enacted at both federal and state levels starting in late 2008 and early 2009.

This discussion is focused on a critical evaluation of the relative success rate of the federal Home Affordable Modification Program (HAMP) from March 2009 when the program took effect through 2016 when the program was scheduled to wind down (excluding modifications still in progress; U.S. Treasury, *Making Home Affordable* Q42016). HAMP, whose stated goal is "to offer homeowners who are at risk of foreclosure reduced monthly mortgage payments that are affordable and sustainable over the long-term," has functioned as a voluntary program that relies upon loan servicers to modify the loans of struggling homeowners through lower monthly payments, thus lowering the risk of foreclosure (U.S. Treasury Dept. *Making Home Affordable* 2012).

Despite HAMP's status as the largest of the government mortgage modification programs, the program lead to trial {conditional} modifications for just over 2.5 million borrowers at risk of foreclosure by year-end 2016, the overall success rate as measured by the number of permanent modifications relative to total trial modifications initiated appears to have fallen short of the program's potential.

The central argument made here is that the design of the program, which established rather strict criteria for borrowers to be considered for a modification, and relied on the voluntary participation of lenders/servicers, essentially ensured that many borrowers in need of loan modifications would simply not qualify, limiting participation from the outset. In the years immediately following the banking and financial crisis, lending overall slowed significantly, and servicer resistance to participation in mortgage modification efforts was evident in the comparatively small number of actual loans modified. Despite one of the key requirements established under the Temporary Asset Relief Program – that banks' receiving temporary assistance - must *consider* homeowners for a loan modification, there remained no mandatory requirement that lenders actually modify the loans of homeowners at risk of default and foreclosure. This essentially enabled servicers for the most part to decide which borrowers they would work with, typically those viewed as posing the lowest risk of re-default. Even with the modest servicer incentives introduced later in the program, total permanent modifications as a share of all trial modifications initiated was 38.3 percent nationally as of year-end 2016.

An overview of the structure of the Home Affordable Modification Program, not only provides a framework within which to understand the numerous obstacles facing homeowners seeking modifications, but also reveals the underlying rationale of lenders/servicers in the context of these obstacles.

In early 2009, in an effort to reach growing numbers of troubled borrowers, new foreclosure prevention measures were introduced, including the Home Affordable Modification Program (HAMP), the Home Affordable Refinance Program (HARP), and 2MP, a program that offered either modification of or extinguishment of second liens for homeowners who had already refinanced their primary loan under HAMP. In February 2009, the U.S. Treasury Department allocated $50 million in TARP funds to help homeowners struggling with their mortgages.

HAMP and HARP were created as part of the Homeowner Affordability and Stability Plan in an effort to help struggling homeowners avoid foreclosure either by modifying or refinancing their first mortgages. Unlike earlier initiatives such as Help for Homeowners (H4H)[1] that relied entirely upon the voluntary participation of lenders and servicers, HAMP required all banks and lending institutions that received government assistance under the Troubled Asset Relief Program to initiate loan modifications for loans that were eligible under HAMP guidelines (Robinson, 2009). However, while this mandate did not apply to non-TARP banks, it was short-lived. In April of 2009, the Treasury Department stipulated that Help for Homeowners (H4H) should be the primary source for homeowners seeking a modification before applying under HAMP (Robinson, 2009). In essence, the TARP mandate to lender participation was further weakened and the process became more bureaucratic for homeowners. In October 2011, the Federal Housing Finance Agency (FHFA), Fannie Mae, and Freddie Mac improved upon HARP-eligible mortgages by making refinancing possible for borrowers who owed more on their mortgages than their homes were worth (Fannie Mae, 2012).

A further change made to the original HAMP program by the Treasury Department was a second lien modification program, known as 2MP. Participation in 2MP would be limited to borrowers who obtained a second lien on or before January 1, 2009 and who had already attained a first lien modification under HAMP. Under 2MP, a second lien that met these requirements would be eligible for either modification or extinguishment (Robinson, 2009). However, 2MP appeared to have had limited success in attracting lender participation, with an estimated 163,000 second lien modifications from the program's inception through December 2016 (U.S. Treasury Dept., December 2016). The principal obstacle was that a second lien modification was not possible unless the borrower had first obtained a first lien modification. The significant numbers of homeowners in negative equity and the continued decline in home values posed a major obstacle to gaining access to the second lien modification program. This issue is inextricably connected to the problem of voluntary lender/servicer participation.

Further, HAMP's own guidelines, established a number of requirements stipulating which borrowers could qualify for a HAMP modification. From the program's beginnings in March 2009, borrowers with conventional loans would qualify if they were delinquent 60 or more days on their mortgages. This essentially disqualified hundreds of thousands of borrowers (if not more), particularly those who held subprime loans and were not yet delinquent. These borrowers were among the first to feel the impact of the housing market collapse, and unaffordable mortgages as a result of spiraling interest rates that were built into their initial loan terms. At the same time, conventional loan holders who may have suffered a recent job or income loss, but otherwise had been in good standing on their loan payments were required to be in delinquency to even begin the process of qualifying for a modification that might either reduce their principal balance or their interest rate.

Only in late January 2012 did the Treasury Department announce changes that expanded eligibility to borrowers with non-Government Sponsored Enterprise (GSE) loans (a move that benefited many subprime loan holders), established more flexible debt-income criteria, and allowed non-owner-occupied properties (i.e. those being rented or vacant properties which were being offered for rent) to qualify for a HAMP loan modification. These changes went into effect on June 1, 2012 (U.S. Treasury, 2012).

Nevertheless, it is argued here that the voluntary feature of the government mortgage loan modification programs, and of HAMP in particular, coupled with the somewhat stringent requirements for borrowers to even qualify for a HAMP trial modification during the first four years of the program, was a significant barrier to the achievement of a higher permanent modification success rate and ultimately reduced its impact on the nation's rising foreclosure rate.

This voluntary feature, which is closely tied to the workings of mortgage markets in which investor interests are prioritized, largely limited the program's success rate. There is strong evidence in the literature that this priority took the form of weighing the financial costs of foreclosure versus modification. The findings suggest that foreclosure often resulted not in a reduction of losses but rather increased losses for investors.

Thus, the HAMP program essentially deferred to the principal objectives of lenders and servicers to consider the long-term outcomes for investors of any modification activities.

To obtain further insight into the factors that may have accounted for the program's diminished impact, this inquiry begins with an examination of several variables that have been frequently cited in the literature as constraints on the ability of borrowers to qualify for a loan modification. These challenges include the large number of homes with negative equity, the high-risk nature of many loans, particularly in the subprime market, the type of loan modification received, and the complications posed by the initial requirement that a first lien modification is conditional on the ability to obtain a second lien modification. In many cases, the lender/servicer of the second lien was different from the servicer of the first lien, which complicated the modification process. These challenges are then considered in the light of the literature on the character of mortgage markets which accorded priority to reducing investor losses and which ultimately shaped the decisions of lenders and servicers to refrain from engaging in larger scale modification efforts.

Within the context of a voluntary modification program, the kind of analysis and policy that informed actions taken by lenders in the interest of investors on failing mortgages appears to have clearly resulted in a significantly higher than socially optimal rate of foreclosure. It is apparent that both foreclosure and modification are costly. However, there are clearly differing circumstances in which modification would be the less costly course of action and where foreclosure is cost saving. Is there, from an economic cost-benefit perspective, a rate of foreclosure that is acceptable from the perspective of lenders and investors – where the costs of foreclosure are essentially equal to the costs of preventing them through a loan modification? Under a policy *requiring* lenders/servicers to evaluate applicants for modification, HAMP could have realized a higher rate of success if such parameters had been in place. Thus, this study poses a theoretical model that offers an alternative methodology for assessing the profitability of foreclosure versus modification based upon a framework for more systematically determining where modification can be cost saving for the investor as an alternative to foreclosure. At the same time, the theory considers the particular social and economic circumstances in which modification offers greater benefits for the borrower, the community and the economy overall.

2. REVIEW OF THE LITERATURE: OBSTACLES TO LOAN MODIFICATION

A number of factors have been widely identified in the literature that would appear to pose challenges to the success of loan modification programs. The assessment in much of the literature is that the forces that led to the foreclosure crisis continued to pose barriers to successful loan modification. These factors include the pervasive practice of low doc and no doc loans along with poor underwriting standards and deteriorating loan quality, particularly in the subprime market (Been, et al., 2011; LaCour-Little, et al., 2009); the large numbers of mortgages with second liens (Been, et al, 2011; LaCour-Little, et al., 2009); smaller down payments and a run-up in borrowing against home equity while home prices were still rising, coupled with declines in home price appreciation that began well before the crisis unfolded (Gerardi, et al, 2011).

This literature considers these dynamics, along with rising and persistent high unemployment in the wake of the crisis, the worsening negative equity position of a growing number of borrowers, and stagnant and/or continued weak recovery in home values as posing significant challenges to attaining higher rates of successful modification.

LaCour-Little, et al (2013), studying a sample of 218,000 ALT-A and subprime home loans originated between 2000 and 2007 and securitized by Bear Stearns, found that the share with full documentation declined quite significantly from 42.4 percent to 21.6 percent over this period, while the share with low documentation increased from 11.5 percent to 69.3 percent. The loans, close to three-quarters of which were ARMs, were found to be associated with a significantly high default risk (2013).

The large number of mortgages with second liens is identified as posing one of the greatest impediments to refinancing (Been, et. al., 2011; LaCour-Little, et. Al. 2009). It is estimated that between 40 and 45 percent of new mortgage loans originated at the height of the housing boom (2005-2007) included a second lien or piggyback mortgage which enabled borrowers with less than a 20 percent down payment to purchase a home, particularly in high cost coastal markets and in markets where house prices accelerated comparatively rapidly (Haughwout, et. al., 2012). This same research also documents that both the number of and value of closed end second liens as opposed to home equity lines of credit (HELOCs) constituted a relatively small percentage of originations in 1999 compared with their peak in 2006.

A broad cross-section of the literature is largely consistent in pinpointing the crisis in the subprime mortgage market beginning in 2006 as the catalyst for much of the larger housing market collapse that followed (Gerardi, and Willen, 2008; Gerardi, et al., 2011; Been, et al., 2011; Rugh and Massey, 2010; Bromley, et al., 2008).

In the years immediately prior to the housing market collapse, increasing numbers of borrowers, particularly in the subprime market, were making very small down payments at the time of purchase, and in many cases, putting zero money down. At the same time, many borrowers who had purchased years before the onset of the crisis, had been withdrawing extraordinary amounts of equity while home prices were still rising, (Gerardi, et al, 2011). This created heightened risk once home prices stalled and began their steep decline. These two conditions alone would clearly pose challenges to refinancing in a down market. After the market peaked, large numbers of homeowners – both subprime and prime - found themselves with negative equity.

Other studies examining the mixed success rates of mortgage modification efforts focus on the persistent complications posed by second liens, negative equity, the failure of modifications to reduce principal balances, and the perception that modification poses a relatively greater cost to investors than foreclosure. Also considered are the shortcomings of the various government loan modification programs introduced in the wake of the foreclosure crisis.

Been, et. al. (2011) point out that HAMP's success was to a significant degree constrained by the presence of a second mortgage. "Second liens significantly complicate modifications because first lien

holders may lose their senior status upon modification," and thus first lien holders are reluctant to agree to participate in a modification unless second lien holders agree to subordinate their liens to the newly modified mortgage (pg 382). As the authors point out, few have chosen to do so. Examining a sample of zip code-level and state data, LaCour-Little, et al., (2009), found that the percentage of piggyback originations from 2001 – 2008 was positively correlated with higher foreclosure rates in subsequent years. Their findings confirm that second liens rose rapidly during the housing boom and were a major contributing factor to underwater mortgages in the face of the sharp decline of home prices after the peak. They specifically looked at whether states and zip codes with a higher proportion of piggyback loans originated during the 2001 - 2006 period experienced increased rates of delinquency and foreclosure. Their findings indicated that second lien originations to subprime borrowers were significantly related to higher rates of foreclosure after 2006. This outcome strongly suggests that borrowers with second liens were likely to be less successful in obtaining a loan modification. However, the findings did not especially hold for *prime* second-lien borrowers (LaCour-Little, et al, 2009). Nevertheless, given the time of their study, it may have been too early to see the full effects of declining home equity coupled with a second lien, which affected large swaths of the home-owning population nationwide, as unemployment rose and home prices continued to decline throughout 2010 and 2011.

The problem of rapidly deteriorating home equity as housing prices fell posed another hurdle to borrowers hoping to qualify for a loan modification. Not until the introduction of the Home Affordable Refinance Program's 'HARP 2.0' in 2011, which allowed refinancing of up to 125 percent of a home's original mortgage, was the problem of rising negative equity as a barrier to qualifying for a loan modification directly addressed (U.S. Treasury, 2011).

Other inquiries, conducted relatively early in the course of the rapid rise in distressed properties, found strong evidence that principal balance reductions were associated with the strongest modification success rates for borrowers. The State Foreclosure Prevention Working Group(SFPWG)[2] (Aug. 2010), analyzing a longitudinal dataset of nine loan servicers in New York State in 2007 before the crisis peaked, found that modifications that included significant reductions in the principal balance tended to have lower re-default rates than their counterparts. This finding led the group to recommend reducing principal balances on loans in areas experiencing significant home price declines. However, modifications with a significant reduction in principal balance represented just 20 percent of the loan modifications that the State Foreclosure Prevention Working Group (SFPWG, 2010) studied.

Similarly, Querci and Ding (2009) found that borrowers were less likely to re-default on their home mortgage when their monthly payments were reduced through a balance-reducing loan modification. Using data from a large sample of recently modified subprime loans, the authors looked at the question of why some loan modifications were more likely to re-default than others. At the same time, they examined the characteristics of modifications that were more likely to re-default within a short-term period. Their findings confirmed that modifications that involved a significant reduction in mortgage payments tended to result in

more sustainable short-term modifications, and that re-default rates are further reduced when payment reductions also include a reduction in principal balances.[3] Nevertheless, such modifications were often the exception as reflected in the 2010 SFPWG study.

With the onset of the financial crisis in late 2008, the SFPWG concluded that a comprehensive approach to loan modification was necessary. At the time the organization issued its fourth report in January 2010, it was estimated that just four out of ten seriously delinquent borrowers were on track for any kind of loan modification. The authors also concluded that while the HAMP program increased the percentage of borrowers participating in some form of loan modification, the rapidly rising number of such delinquent borrowers meant that HAMP had merely been able to slow the foreclosure crisis, and that its efforts have not been able to keep pace with the rising scale of delinquencies (SFPWG, Jan. 2010).

However, despite the compelling evidence that HAMP was at best able to slow the pace of the foreclosure process by gradually qualifying more borrowers for modifications, and the findings of studies that principal balance reductions were clearly most successful in reducing re-default risk and benefitting borrowers, modifications continued to proceed at a relatively slow pace relative to the rising rates of default and foreclosure (SFPWG, 2010; Statistic Brain, 2016) and the share of such modifications that reduced principal balances remained comparatively small.

The numerous obstacles to successful loan modification for countless borrowers in the aftermath of the housing crisis that these studies reveal appear to be linked to a key issue: that the voluntary design of the HAMP program, together with the primary goal of lenders/servicers to prioritize efforts to protect investors – a goal which itself shaped their voluntary participation – posed a significant barrier to a more robust success rate for HAMP. The voluntary structure of the HAMP program in essence deferred to the principal objective of lenders and servicers to consider the long-term outcomes for investors of any modification activities. This is highlighted in several studies that have placed the lower than potential rate of modification in perspective.

Foote et. al., (2010); Adelino, et. al., (2009), White, (2009) and Piekorski (2011) focus on the central issue of potential losses to investors of re-default risk of modified loans in the face of rising job loss and home price depreciation. Foote et. al. (2010) find evidence that the unwillingness of many mortgage servicers to make large scale modifications is linked to the finding that the losses to investors from foreclosure are actually less than from modification, especially when modifications are done ‘en masse’ (2010). This would seem counterintuitive. However, the authors provide evidence that the added risk of borrowers re-defaulting on the modified loan enhances the potential losses to investors from modification vs. foreclosure. Thus, they conclude that foreclosure prevention policies aimed at reducing high debt-to-income ratios and borrowers’ interest rates may not effectively reduce what they point to as the key source of loan defaults – falling home prices and job loss (pg. 91). In other words, perceived re-default risk may have much more to do with plummeting values of the asset – homes -combined with rapidly escalating unemployment across the economy, both of which pose a high re-default risk. From a net present value

perspective, they argue that most "potential modifications are negative-NPV transactions from the standpoint of investors" (120).

Finally, there is the argument that some servicers were better equipped than others by virtue of their organizational capacity to process larger numbers of modifications than were others resulting in uneven outcomes across servicers (Agarwal et. al. (2012). Thus, the success rate of mortgage debt adjustment succeeded in assisting a significantly smaller percentage of households – approximately 30 percent based on the authors' study - relative to the total who qualified (pg. 4). The authors examine the variance in lender participation in the context of NPV considerations (pg.24). They find that a number of modified loans re-default following modification, while others that were initially delinquent emerge from delinquency without modification (i.e. they are self-cured). Overall, they find that a loan that is delinquent and which does not 'self-cure' has a 50 to 60 percent probability of ending up in foreclosure (Argarwal et. al., 2012).

What many of these studies share in common, including those that have analyzed mortgage loan data sets, (Adelino, et. al., 2009, Foote, et. al., 2010, LaCour-Little, 2009) is the conclusion that the securitization process in which mortgage loans are re-sold as investments, was not responsible for the low rate of modifications. Rather, the role of NPV calculations in deciding whether losses from foreclosure will be less than those from modification is central to the decision of a lender/servicer to participate in modifying a loan. The studies reach similar conclusions that confirm the centrality of loss mitigation concerns and therefore offer added insight into the problems associated with voluntary participation in HAMP.

Thus, the many obvious obstacles to obtaining a loan modification discussed in the literature on negative equity, second liens, and loan modification type, can be understood in the larger context of the very risk that modification posed to investors and that lenders/servicers weighed in considering the extent of their participation in modification efforts. Investors were clearly aware of the growing risk posed by rising and stubbornly high unemployment coupled with the deteriorating value of their assets – homes - and their central concern which was to protect the value of that asset. This perceived risk illuminates the pervasive uncertainty about the future direction of the economy that constrained broader participation in HAMP and provides a more far-reaching context within which to understand the program's limited success.

3. SUMMARY OF HAMP OUTCOMES

An overview of key program outcomes between 2009 through 2011, as housing prices showed signs of reaching a bottom, offers some perspective on the overwhelming hurdles HAMP faced given its built-in constraints as well as the many challenges borrowers confronted in renegotiating their home mortgages. This may place into perspective the risks borrowers were considered to pose in the context of depressed home prices and worsening economic conditions. In 2011 IIQ, 22.1 percent of residential properties with a mortgage - an estimated 10.7 million homes – were still in negative equity nationwide, while more than two-thirds of mortgage holders on such properties were paying above market interest rates (CoreLogic, 2011). This reflected only slight improvement from 2009 IIIQ when 24 percent of properties were in negative equity

(Warren, 2009; CoreLogic, Nov. 2009). That the negative equity rate remained essentially flat for two years following the official end of the recession in June 2009 reveals one of the challenges faced by HAMP as both modification applications and foreclosures also continued to rise.

By year-end 2016, the percentage of mortgaged homes in negative equity declined significantly to 6.2 percent. Interestingly, however, the overwhelming majority of homes with equity at year end 2016 – 96 percent - were concentrated at the higher end of the market, where homes are valued at $200,000 and over (Core Logic, IVQ 2016).

The negative equity problem was also reflected in home price declines during the crisis and in the years immediately following. The seasonally adjusted S&P Case-Shiller 20-city U.S. national home price index (quarterly) shows that the national market bottomed out in 2012 IQ after peaking in 2006 IIIQ. Over this period, U.S. home prices overall lost 36.9 percent of their value (S&P Case Shiller, 2017). By 2016 IVQ, home prices recovered 27 percent from their pre-recession peak.

The unemployment rate, which peaked at 10.0 percent in October 2009, did not dip below 6 percent until September 2014 (U.S. Bureau of Labor Statistics).

Nationally, 962,209 homeowners out of 2,511,344 who entered into a first lien trial modification from HAMP's inception in March 2009 through December 2016 completed a permanent modification of their home mortgage through the program. In total, 1,683,112 borrowers entered into a trial first lien permanent modification. (U.S. Department of the Treasury, QIV 2016). Among completed modifications, this represents a national success rate of 38.3 percent and 67 percent when cumulative trial modifications are included. 200,552 permanent modifications featured a principal balance reduction through 2016, while 290,279 such modifications were in trial accounting for just 8.0 percent and 11.6 percent, respectively of all first lien trial modifications started.[4] Under HAMP's Home Affordable Foreclosure Alternative, borrowers denied a HAMP modification were required to be considered for a plan enabling them to exit their mortgage obligation through either a deed-in-lieu or a short sale. However, individual investors could impose further eligibility requirements (U.S. Department of the Treasury, QIV 2016) which could have made obtaining this option more of a challenge for some borrowers.

Through December 2016, successful second lien modifications also represented a relatively small proportion of total modifications nationally. As of year-end 2016, 163,140 second liens had entered the 2MP modification program. Just 48,318 of these resulted in a full extinguishment of the second lien, while another 10, 470 received a partial lien extinguishment. The remaining 79,343 second liens were in active modification status. Just five servicers accounted for 85 percent of second lien modifications through December 2016 (U.S. Treasury, 2016), representing a fraction of total servicers nationwide.

A further comparison with national foreclosure data indicates an even smaller successful modification rate. More than 6.2 million foreclosures nationally were completed from 2009 through 2016 (Core Logic, March 2017). When compared to the 962,209 distressed mortgage holders who received a permanent loan modification, the percentage of successful modifications drops to 15.5 percent of troubled mortgages over

this seven-year period. The data on completed foreclosures appears to confirm that foreclosure was the overwhelming direction taken relative to modifications.

The percent share of loan modifications among the top seven servicers through December 2016 reveals mixed outcomes. These ranged from a low of 2.7 percent for CitiMortgage, Inc. to a high of just 21 percent for Ocwen Loan Servicing, LLC, a subprime lender. Ocwen also accounted for the largest percentage of modifications featuring a principal balance reduction at nearly 49 percent of the total. These data also highlight the comparatively low percentage – 10.2 percent - that principal balance reductions made up of total modifications (U.S. Treasury, 2016).

HAMP modifications among investor groups include all HAMP Tiers 1 and 2 and Streamline permanent modifications. The data suggest that loans held by the GSEs and by private investors had a larger proportion of permanent modifications compared to loans held in portfolio. The GSEs accounted for 39 percent and private investors 44.1 percent of permanent modifications relative to just 16.9 percent for loans held in portfolio.

The comparatively small number of permanent modifications attained relative to foreclosures from the inception of the Home Affordable Modification Program in early 2009 through year-end 2016 should also be viewed from the perspective of the impact of foreclosures in an economy that experienced a prolonged and steep decline following the housing market collapse. Foreclosures accelerated rapidly from 2007 through 2010, and while the policy response in the form of modification initiatives such as Help for Homeowners and HAMP helped to reduce the impact, the sizeable number of foreclosed properties in communities across the country had a direct impact on already declining property values in those markets. If a servicer or investor is more reluctant to modify a distressed underwater mortgage, the choice to foreclose instead simply exacerbates the problem and adds to a glut of vacated or abandoned properties, further dragging down the value of surrounding homes, increasing the economic costs. This affects all homeowners, not just those struggling to pay their mortgages and as the repercussions in the form of declining household wealth are experienced across the economy, the social and economic costs of foreclosure are magnified.

The following discussion outlines a theoretical framework for understanding the economic issues and pressures that HAMP intended to address in the midst of a foreclosure crisis in which the economic interests of borrowers, servicers and investors were often at odds. The policy structure of the program ultimately resulted in a lower than optimal modification success rate. Given that foreclosures are both costly to prevent and to carry out, the discussion proposes a model for how the modification success rate could have been greater in the context of HAMP's voluntary structure and how the problem of lenders who made risky mortgages that contributed to a large share of the problem might have been addressed differently.

Table 1: Making Home Affordable Program Activity by Servicer: March 2009 - December 2016

Servicer	HAMP Tier 1 Permanent Modifications	HAMP Tier 2 Permanent Modifications	Streamline HAMP Permanent Modifications	PRA Permanent Modifications	2MP Modifications	HAFA[12] non-GSE Transactions Completed	Servicer Total
Bank of America, N.A.	103,134	8,004	N/A*	5,885	38,499	49,861	205,383
CitiMortgage, Inc.	32,881	3,784	0	3,233	20,341	2,487	62,726
JPMorgan Chase Bank, N.A.	162,915	5,579	2,077	25,441	44,703	38,215	278,860
Nationstar Mortgage LLC	183,837	25,425	1,580	11,225	9,810	11,270	243,147
Ocwen Loan Servicing, LLC	244,433	80,809	13,951	115,433	N/A*	29,128	483,754
Select Portfolio Servicing, Inc.	114,438	27,565	8,503	21,360	N/A*	22,374	194,240
Wells Fargo Bank, N.A.	199,734	12,368	0	30,432	25,032	44,639	312,205
Other Servicers	417,151	33,495	1,449	22,572	24,755	30,982	530,404
Total	**1,458,523**	**197,029**	**27,560**	**235,581**	**163,140**	**228,956**	**2,310,719**

Source: U.S. Treasury Department, December 2016 MHA Report. (It should be noted here that servicers report all trial modifications as permanent modifications).
*Servicer does not participate in either Streamline HAMP or HAMP 2MP.
[11] Principal Reduction Alternative
[12] HAFA: Home Affordable Foreclosure Alternative (This program offered incentives to homeowners to terminate their mortgage commitment or to sell a rental property through a short sales or a deed-in-lieu of foreclosure).
Notes from "Select HAMP Modification Characteristics" (p. 8, Making Home Affordable: HAMP Program Results: Program Performance Report Fourth Quarter 2016)
**Under HAMP Tier 1, servicers apply the modification steps in sequence until the homeowner's post-modification front-end debt-to-income (DTI) ratio is 31%. The impact of each modification step can vary to achieve the target of 31%.
**Under HAMP Tier 2, servicers apply the modification steps simultaneously to achieve a post-modification DTI that falls within an allowable range (subject to investor restrictions). HAMP Tier 2 applies to non-GSE mortgages only.
**Under Streamline HAMP, seriously delinquent homeowners who have not been able to complete a HAMP application may be eligible to receive mortgage assistance through a combination of modification steps similar to HAMP Tier 2. Unlike Tier 1 and Tier 2, Streamline HAMP does not require that borrowers document their income.

Table 2: HAMP Permanent Modifications by Investor

Servicer	All HAMP Permanent Modifications			
	GSE	**Private**	**Portfolio**	**Total**
Bank of America, N.A.	39,182	53,663	18,293	111,138
CitiMortgage, Inc.	15,182	9,223	12,260	36,665
JPMorgan Chase Bank, N.A.	69,483	56,806	44,282	170,571
Nationstar Mortgage LLC	119,528	82,379	8,935	210,842
Ocwen Loan Servicing, LLC	23,257	293,966	21,970	339,193
Select Portfolio Servicing, Inc.	14,654	115,137	20,715	150,506
Wells Fargo Bank, N.A.	80,487	42,034	89,581	212,102
Other Servicers	295,093	89,342	67,660	452,095
Total	**656,866**	**742,550**	**283,696**	**1,683,112**

Source: U.S. Treasury Department, December 2016 MHA Report (all permanent modifications started are reported by servicers as permanent HAMP modifications).

4. UNDERSTANDING THE ECONOMIC CONFLICTS: A THEORETICAL MODEL

What were the economic issues at the heart of the conflicting interests of borrowers and lenders/servicers and investors? How did the tensions between these competing economic interests result in a less than optimal rate of successful loan modification and a higher rate of foreclosures? How did the structure of HAMP policy contribute to a less than optimal loan modification rate? How might outcomes have been different had policy been structured with the goal of bringing the economic interests of borrowers as well as servicers/investors into closer alignment?

To place the economic issues in perspective, consider the developments that contributed to the crisis. Banks and other lenders, encouraged by an improving economy in the early 2000's and an increased demand for homes made more mortgage loans to buyers. At the same time subprime lending, once a very small portion of the mortgage market grew to constitute 20 percent of all loans at the peak of the housing boom by 2005 from less than 5 percent in 1994 (Doms, et.al.,2007). An expanding range of increasingly exotic mortgages enabled buyers – both prime and subprime – to purchase homes that they otherwise may not have been able to afford. Subprime loans which typically came with a higher rate of interest were particularly attractive to investors in mortgage backed securities as these offered a higher rate of return. The combination of exotic, risky and high interest loans and the increasing demand from investors for more such loans encouraged more risk taking on the part of lenders as more borrowers, including many who would otherwise not qualify were approved for loans, often with little documentation.

As borrower delinquency rates, initially on subprime loans, began to noticeably rise during 2006 into 2007, it was initially believed that the crisis could be contained within the subprime sector but as

delinquencies trended upward even among seemingly solid borrowers with fixed interest loans, it became evident that it could not. With the deteriorating economy shedding jobs at accelerated pace, the foreclosure rate increased rapidly, pointing to the mounting social and economic costs to come. As the housing market collapse unfolded and a full blown financial crisis now threatened the entire economy, investors sought to minimize their losses and many borrowers sought to hold on to their homes. The tension between these competing interests intensified as foreclosure prevention measures were implemented in a number of states. New York, for instance, mandated pre-foreclosure notices to delinquent borrowers and set up housing counseling and other services to assist borrowers at risk of losing their homes. It became clear that a torrent of foreclosed properties hurts not just families, but the viability of entire communities, local property tax revenue, and the entire economy. From the investors' perspective, the overriding interest in protecting the value of their assets and cutting their losses in the face of declining home values left little motivation to modify loans.

The policy response - the Home Affordable Modification Program - seemed to recognize these economic tensions. However, in seeking to align the interests of borrowers and lenders/servicers and the investors they represent, the program fell far short. A higher rate of successful modifications would have been likely if certain program features had not limited the scope of both lender and borrower participation by establishing unrealistic barriers. The requirement that any applicant must already be delinquent on payments by 60 days or more essentially meant that many borrowers in high-cost mortgages had to choose between a greater likelihood of default sometime in the future or deliberately not making mortgage payments and gambling that they could qualify for a modification. Second, had there been no requirement that second lien modifications could only be considered if a first lien modification had first been attained, many more borrowers could have qualified. On the lender/servicer side, the requirement that borrowers be considered for a modification should have applied to *all* lenders, not just those receiving TARP funds. Further, had non-GSE lenders, many of whom were originators of subprime loans, been included in that requirement from the inception of HAMP, a larger pool of loans would have qualified for modification. All of these constraints resulted in far fewer permanent modifications than otherwise would have been possible. A 38.3 percent success rate among more than 2.5 million applicants who qualified and a 15.5 percent rate relative to 6.2 million foreclosures completed over this period is far less than optimal.

Beyond these shortcomings, what might have produced a more effective policy outcome? From the perspective of servicers and the investors holding mortgages, is there a rate of foreclosure where the harm resulting from foreclosures is roughly equivalent to the cost of averting them? Given that the economic costs of foreclosure extend well beyond the impact on the individual borrower and the individual servicer/investor, how could foreclosure rates have been reduced (and modification rates increased) to the point where the social and economic costs associated with foreclosures were in closer alignment with servicer/investor costs of preventing them? The following discussion proposes measures that could be a step in the direction of narrowing that gap.

First, suppose HAMP guidelines had *mandated* that all lenders/servicers participate in the program, considering the cost savings from modification vs. foreclosure based on a cost benefit approach - evaluating the actual costs of each decision. Lender/servicer motivations for foreclosing rather than modifying would be driven by an estimate of the actual costs of carrying out each action. So how can the costs be placed in context? An estimated 11.3 million – 24 percent – of homes with a mortgage were in negative equity in Q3 2009 in the depths of the Great Recession, while the average dollar amount of negative equity at the time was $70,000 (CoreLogic Q3 2009). The higher the negative equity share, the greater was the probability of receiving a pre-foreclosure notice (CoreLogic Q3 2009) and thus, the greater the probability of foreclosure. It is quite likely, given the 6.2 million foreclosures between 2009 and 2016, that many of these properties ended in foreclosure.

1) Assume that the original mortgage = M_0 and the modified mortgage with principal balance = M_1 . The costs (C) of modification equal total principal balance reduced of the negative equity amount (N).

$$C = M_0 - N = M_1.$$

Assume that M_0 = $300, 000 and negative equity = $70,000; then M_1 ≥ $230,000, but ≤ $299,999.

If it is assumed that the average household mortgage was $300,000 in 2009 with average negative equity of $70,000 for the 24 percent of homes with a mortgage this constitutes a loss for the homeowner and the lender, either of whom could be on the losing end as a result. The borrower who can afford to, will continue making the mortgage payments. However, in the case of borrowers who put little money down and purchased a house that their incomes could not support – the question is how much of a mortgage they can afford to pay. If that number lies somewhere between $230,000 and $299,999, the lender (or investor holding the mortgage), should have an incentive to negotiate a principal balance reduction. A new mortgage of $230,000 would mean that both borrower and lender break even since the home is now worth $230,000.

2) Now suppose the borrower can afford a modified mortgage with a principal balance reduction greater than $230,000. The borrower is assuming some of the negative equity. Here N assumes a value of $50,000. In this case, if a principal balance reduction to $250,000 is negotiated between borrower and lender, the lender attains an asset whose value is $230,000 and acquires $20,000, bypassing the expense of a potentially costly foreclosure process. Assuming half of the 11.3 million mortgage holders in negative equity were among those who could afford to pay this new principal balance, 5.65 million foreclosures would be avoided, borrowers would not lose the money already invested in their homes, and investors would retain their assets.

3) Suppose another 3 million borrowers could not afford a mortgage of $250,000, but could manage a loan payment between $200,000 and $230,000, then the costs and benefits of foreclosure – such as the legal costs of carrying out foreclosure actions against borrowers, continued erosion of home values, the costs of preparing documentation, etc. must be weighed, since any renegotiated mortgage less than the property's value would impose a cost on the lender. Assuming these foreclosure costs average $30,000,

the lender could agree to reduce the principal balance to $215,000 and potentially save $15,000 in foreclosure costs. Here, the costs of foreclosure avoidance (F) are still positive and:

$$M_1 = M_0 - N - (30{,}000 - 15{,}000) = \$215{,}000.$$

The balance could be further reduced to as low as $200,000 for some borrowers and the lender would break even. Overall, such a scenario would potentially avoid another 3 million foreclosures and investors would again retain their assets.

$$M_1 = M_0 - N - (30{,}000 - 0) = \$200{,}000.$$

What about the cost to investors? While ultimately, the costs of modification are borne by the investor while benefiting the homeowner, additional requirements stipulating that lender and investor share in the gain from a modified mortgage would ensure the benefits are distributed to both parties. In the case where the lender/servicer gains $20,000 from a modified loan of $250,000, half of that acquisition would go to the investor. In the case of the $215,000 loan, the same requirement would apply. Assuming some modifications result in a principal balance reduction to $200,000, the investor writes off the loss of $30,000 in equity at the time the mortgage is modified, but still retains the asset. At the same time, the larger economy benefits from having fewer foreclosed houses on the market to further bring down property values, drain tax revenues and further weaken economic recovery.

However, what about borrowers who are still unable to afford a $200,000 modified mortgage? Employing the same example, in such cases where foreclosure is more likely, the lender would acquire the property and find a new buyer. However, several questions arise. What kind of loan was made to the borrower (i.e., high interest, interest only, negative amortizing, etc.)? Was little or no documentation required? If the loan was high risk, the lender should be held accountable for contributing to the excessive systemic risk that led to widespread default rates and the plummeting home values, job losses and financial crisis that nearly led to economic collapse.

Originators of high-risk loans, many of which were subprime, would be required to bear some of the costs of their decisions. A number of factors might be weighed in determining that cost. In the case of borrowers who put money down, how much did they pay? How much principal had already been paid? Were any improvements made to the home? In every instance, what is the cost to the homeowner of packing up and moving out? Compensation to displaced homeowners should at the very least be based on such costs borne by the homeowner. Further, the lender would be required to contribute a portion of those costs to state and/or federal level foreclosure prevention programs. If it is assumed that homeowner costs are $10,000, then the lender pays $8,000 to the borrower to offset the costs of finding new housing and contributes $2,000 to the state or federal program. Such measures send a clear message: that contributing to systemic risk requires sharing responsibility and payment of some kind of penalty.

Finally, there were clearly many otherwise creditworthy borrowers who became delinquent and went into default following job loss. As the unemployment rate quickly rose during the Great Recession, the odds that these homeowners would be able to resume payments after just a few months and/or upon finding a

new job were slim. Average unemployment duration was more than 24 weeks in 2009, and averaged 37 weeks over 2010-2012 when the effects of the recession were still being felt (Statista, 2017), making default far more likely.

HAMP included an unemployment program in which homeowners could be approved either for a forbearance plan with some payment required or with no payment required for 12 months, allowing homeowners to seek new employment without losing their homes. However, of the 46,485 applicants who were approved for and started the plans, 24 percent remained current on payments after 12 months (MHA Quarterly Report, Q4 2016). Clearly, this has some positive economic impact. A 2016 study found that …" foreclosure delay during the recession improved the quality of new employment matches, raised national income by about 0.3 percent and increased homeownership by about 800,000 units (Herkenhoff and Ohanian, 2016).

However, might the success rate have been higher if, in addition to forbearance, unemployed borrowers had been evaluated for a mortgage modification with a principal balance reduction, applying the same guidelines as those detailed above? In this case, once borrowers resume making the resulting lower payments, the end result may have resulted in a lower re-default rate. Given that for another 32 percent of homeowners, the final outcome was bankruptcy, action pending or a charge off, while another 6 percent re-entered the foreclosure process or a deed-in-lieu, principal balance reductions could have produced a stronger forbearance success rate.

5. CONCLUSIONS

Clearly the challenges borrowers faced in renegotiating their home mortgages illustrates the risks they were considered to pose in the context of depressed home prices and worsening economic conditions.

The application of a model based on the framework posed here would bring foreclosure prevention into closer alignment with the goal of protecting investors. This actually *requires* lenders/servicers to reach a modification agreement with applicants where possible. The result, where borrowers, lenders/servicers and investors benefit from the outcome not only reduces the social disruptions caused by massive foreclosures, but helps to minimize the larger economic costs, potentially easing the impact of a steep downturn, stabilizing affected communities, and stemming the blight of foreclosed properties in neighborhoods already experiencing eroding home values. At the same time, the property tax base in those communities is stabilized at a time when revenue needs are greatest.

Preventing even a significant percentage of the 6.2 million foreclosures that occurred between 2009 and 2016 could have resulted in significantly less income and wealth loss in the economy, while avoiding the costs of foreclosure incurred both by the homeowner and financial institutions, as well as the various government entities involved in legal processing of foreclosure actions.

The data and the analysis reviewed on HAMP's outcomes in successful permanent modification of distressed home loans suggest that policy design is critical. National policy design must be more robust in

addressing the larger picture, in this case the economy-wide costs of widespread foreclosure both during and in the immediate aftermath of a steep downturn such as the Great Recession. This requires lender/servicer participation and the establishment of a set of guidelines for their participation. Those posed here offer a framework for thinking about such participation. At the very least, this should oblige lenders/servicers to evaluate not only the costs but the immediate and long-term benefits of loan modification. The perspective offered here may offer a starting point for more formal analyses that evaluate sample data on foreclosure outcomes over the course of HAMP's modification program.

ENDNOTES

1. Hope for Homeowners (H4H), the initial program introduced in late 2008 in the final months of the Bush Administration, enabled underwater borrowers to refinance into an FHA guaranteed mortgage. H4H relied upon the voluntary participation of lenders and servicers. Prior to 2008, when the first signs of soaring foreclosure rates began to appear principally in the subprime market, efforts which encouraged lenders and servicers to work with subprime borrowers to modify their high variable interest rate loans into fixed rate loans, relied upon voluntary participation in such efforts by lenders and loan servicers.
2. The Subprime Foreclosure Prevention Working Group consisted of several state attorneys general and state bank supervisors.
3. Following changes made to the Treasury Department's guidelines under the terms of the HAMP Principal Reduction Alternative, servicers of non-GSE loans were required to evaluate borrowers for a principal reduction (although they are not required to provide such a modification) under the terms of the national mortgage settlement (U.S. Treasury, Dec. 2012) with the nation's five largest servicers. As a result, many servicers began to increase the use of non-PRA principal reductions after 2012.
4. Based on the details of U.S. Treasury HAMP reports, one of the principal reasons has had much to do with policy guidelines and limitations under HAMP that were still in effect through year-end 2012. Those guidelines stated that while both GSE and non-GSE loans (i.e. many subprime loans) were eligible to participate in a HAMP modification, GSE policy (Fannie Mae and Freddie Mac) stipulated that servicers can only offer a principal balance reduction – a PRA (or Principal Reduction Alternative) on non-GSE modifications under HAMP (2012).

REFERENCES

Adelino, Manuel, Kristopher Gerardi, and Paul S. Willen. July 2009. "Why Don't Lenders Renegotiate More Home Mortgages? Redefaults, Self-Cures, and Securitization." Federal Reserve Bank of Boston, *Public Policy Discussion Papers,* 09 (4), 1 – 41.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru. Aug. 2012. "Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program." *NBER Working Paper Series,* (18311), 1 – 60.

Been, V., Chan, S., Ellen, I. G., and Madar, J. R. 2011a. "Decoding the Foreclosure Crisis: Causes, Responses, and Consequences." *Point/Counterpoint, Kenneth A. Couch, ed., Journal of Policy Analysis and Management*, 30(2), 381–400. © 2011 by the Association for Public Policy Analysis and Management

Been, V., Chan, S., Gould, I., and Madar, J. R. 2011b. "Negative Equity, Yes, but Not the Whole Story." *Point/Counterpoint, Kenneth A. Couch, ed. Journal of Policy Analysis and Management*, 30(2), 381–400. © 2011 by the Association for Public Policy Analysis and Management

Board of Governors of the Federal Reserve System. 2014. "Independent Foreclosure Review." Retrieved from http://www.federalreserve.gov/consumerinfo/independent-foreclosure-review.htm

Bromley, C., Campen, J., Nafici, S., Rust, A., Smith, G., Stein, K., and van Kerkhove, B. Mar. 2008. "Paying More for the American Dream: The Subprime Shakeout and its Impact on Lower-Income and Minority Communities." Retrieved from http://www.policyarchive.org/handle/10207/19021

Core Logic. "Negative Equity Report." March 2009. Retrieved from http://www.corelogic.com

CoreLogic. Nov. 2011. "CoreLogic Third Quarter 2011 Negative Equity Data Shows Slight Decline But Remains Elevated." Retrieved from http://www.corelogic.com/about-us/news/corelogic-third-quarter-2011-negative-equity-data-shows-slight-decline-but-remains-elevated.aspx

Core Logic. September 2012. "Negative Equity Report." CORELOGIC® REPORTS NUMBER OF RESIDENTIAL PROPERTIES IN NEGATIVE EQUITY DECREASES AGAIN IN SECOND QUARTER OF 2012. Retrieved from http://www.corelogic.com

Core Logic. December 2012. "National Foreclosure Report." Retrieved from http://www.corelogic.com/research/foreclosure-report/national-foreclosure-report-december-2012.pdf

Core Logic. March 2013. "Negative Equity Report."

Core Logic, Feb. 2013. "Negative Equity Report."

Core Logic, QIV 2016. "Equity Report."

Doms, M., Furlong, F., and Krainer, J. November 2007. "Subprime Mortgage Delinquency Rates." *Federal Reserve Bank of San Francisco* Working paper, 2007-33. Retrieved from http://www.frbsf.org/publications/economics/papers/2007/wp07-33bk.pdf

Duke, Gov. E. A. March 8, 2013. "Comments on Housing and Mortgage Markets." Board of Governors of the Federal Reserve System." Retrieved from **http://www.federalreserve.gov/newsevents/speech/duke20130308a.htm**

Fannie Mae. 2012. "Home Affordable Refinance Program Enhancements." Retrieved on April 2, 2014 from www.eFannieMae.com

Federal Housing Finance Agency. Oct. 2011. "FHFA, Fannie Mae and Freddie Mac Announce HARP Changes to Reach More Borrowers." News release. Retrieved from

Foote, C., Gerardi, K., Goette, L., and Willen, P. April 2010. "Reducing Foreclosures: No Easy Answers." *National Bureau of Economic Research Annual 2009,* 89-138, 24. Retrieved from http://www.nber.org/chapters/c11790.pdf

Gerardi, K. S., and Willen, P. 2008. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods." *Public Policy Discussion Papers*, 08-6, 22. Retrieved from http://www.bos.frb.org/economic/ppdp/2008/ppdp0806.pdf

Gerardi, K., Ross, S.L., and Willen, P. 2011a. "Understanding the Foreclosure Crisis." *Point/Counterpoint*. Kenneth A. Couch, ed., *Journal of Policy Analysis and Management.*30(2), 381–400. © 2011 by the Association for Public Policy Analysis and Management

Gerardi, K., Ross, S.L., and Willen, P. 2011b. "Decoding Misperceptions: The Role of Underwriting and Appropriate Policy Responses." *Point/Counterpoint*. Kenneth A. Couch, ed., *Journal of Policy Analysis and Management.* 30(2): 381–400. © 2011 by the Association for Public Policy Analysis and Management

Gerrity, M. Jan. 17, 2013. "Total U.S. Foreclosures Top 1.8 Million in 2012, a 36 Percent Decline from 2010 Peak." *Realty Trac*. Retrieved from www.realtytrac.com

Haughwout, A., Lee, D. Tracy, J., and van der Klaauw, W. September 2011. "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis". *Federal Reserve Bank of New York Staff Reports*, 514. Retrieved from http://www.newyorkfed.org/research/staff_reports/sr514.pdf

Herkenhoff, K. and Ohanian, L.E. 2016. "Foreclosure Delay and the U.S. Labor Market." Federal Reserve Bank of Minneapolis, Economic Policy Papers, 1.

Humphries, Stan. Aug. 22, 2012. "Negative Equity Declines Slightly on the Back of Modest Home Value Gains," Zillow. Retrieved from: https://www.zillow.com/research/negative-equity-declines-slightly-on-the-back-of-modest-home-value-gains-3134/

LaCour-Little, M., Rosenblatt, E., and Yao, V. 2010. "Home Equity Extraction by Homeowners: 2000-2006." *Journal of Real Estate Research*, 32, 23 – 46.

LaCour-Little, M. and Yang, J., Oct-Dec 2013. "Taking the Lie Out of Liar Loans: The Effect of Reduced Documentation on the Performance and Pricing of Alt-A and Subprime Mortgages," *Journal of Real Estate Research,* 507-554, 35 (4)

New York State Mortgage Settlement. 2012. Retrieved from http://www.nysmortgagesettlement.com/

New York State Mortgage Settlement. 2012. "Payments to Borrowers who Lost their Homes to Foreclosure. 2012." Retrieved from http://www.nationalmortgagesettlement.com/

Quercia, R. and Ding, L. 2009. "Loan Modifications and Re-default Risk: An Examination of Short-term Impacts." *Cityscape: A Journal of Policy Development and Research*. 11(3), 171-193.

Realty Trac. 2012. "August Foreclosure Activity Flat Nationwide, Spikes in Some States - September 11, 2012." Retrieved from www.realtytrac.com

Robinson, Breck. 2009. "An Overview of the Home Affordable Modification Program." *Consumer Compliance Outlook*, third quarter, 2009. *Federal Reserve Bank of Philadelphia*. Retrieved from http://www.philadelphiafed.org/bank-resources/publications/consumer-compliance-outlook/2009/third-quarter/q3_02.cfm

Rugh, J. S. and Massey, D.S., 2010. "Racial Segregation and the American Foreclosure Crisis." *American Sociological Review.* 75(5), 629–651.

S&P CoreLogic Case-Shiller 20-City Composite Home Price NSA Index. Retrieved from https://us.spindices.com/indices/real-estate/sp-corelogic-case-shiller-20-city-composite-home-price-nsa-index

S&P/Case-Shiller U.S. National Home Price Index. Retrieved from https://fred.stlouisfed.org/series/CSUSHPINSA?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=categories

State Foreclosure Prevention Working Group. August 2010. "Redefault Rates Improve for Recent Loan Modifications." *Memorandum on Loan Modification Performance*. Retrieved from http://www.csbs.org/regulatory/Documents/SFPWG/DataReportAug2010.pdf

State Foreclosure Prevention Working Group. January 2010, "Analysis of Mortgage Servicing Performance." Data Report No.4. Retrieved from http://www.csbs.org/regulatory/Documents/SFPWG/DataReportJan2010.pdf

Statista. "Average Duration of Unemployment in the United States from 1990 to 2016 (in weeks)." Retrieved from https://www.statista.com/statistics/217837/average-duration-of-unemployment-in-the-in-the-us/

Statistic Brain, 2016. "Home Foreclosure Statistics." (data compiled from RealtyTrac, Equifax, Federal Reserve). Retrieved from http://www.statisticbrain.com/home-foreclosure-statistics/

Svenja, G. February 20, 2013. "Nearly 2 Million American Homeowners Freed from Negative Equity In 2012." *Zillow Real Estate Research***.** Retrieved from http://www.zillowblog.com/research/2013/02/20/nearly-2-million-american-homeowners-freed-from-negative-equity-in-2012/

U.S. Bureau of Labor Statistics. 2017. Labor Force Statistics from the Current Population Survey. Retrieved from https://data.bls.gov/pdq/SurveyOutputServlet?request_action=wh&graph_name=LN_cpsbref3

U.S. Department of Housing and Urban Development. August 28, 2012. "Housing Scorecard Regional Spotlight Reports: The Success of HARP 2.0." Retrieved from http://harp-mortgage.com/the-success-of-harp-2-0/

U.S. Department of Labor. August 2012. "Employment Situation Summary." Retrieved from http://www.bls.gov/news.release/empsit.t12.htm

U.S. Treasury Department. 2010. "Making Home Affordable, The Home Affordable Modification Program (HAMP), "HAMP Activity by Metropolitan Statistical Area As of January 2010"

U.S. Treasury Department. 2012. "Making Home Affordable, The Home Affordable Modification Program (HAMP)." Retrieved from http://www.treasury.gov

U.S. Treasury Department. 2012. "The FHA's HOPE for Homeowners Program." Retrieved from https://www.treasury.gov/initiatives/financial-stability/TARP-Programs/housing/mha/Pages/hamp.aspx

U.S. Treasury Department. 2011. "The Home Affordable Refinance Program." Retrieved from http://www.treasury.gov

U.S. Treasury Department. December 2012. "December 2012 Making Home Affordable Program Performance Report." Retrieved from http://www.treasury.gov/initiatives/financial-stability/reports/Documents/December%202012%20MHA%20Report%20Final.pdf

U.S. Treasury Department. July 22, 2012. "Making Home Affordable: Home Affordable Modification Program (HAMP)." Retrieved from http://www.treasury.gov/initiatives/financial-stability/TARP-Programs/housing/mha/Pages/hamp.aspx

U.S. Treasury Department. Housing and Urban Development. 2012. "U.S. Housing Market Conditions 4th Quarter 2012" Historical data." Retrieved from http://www.huduser.org/portal/periodicals/ushmc/winter12/index.html

U.S. Treasury Department. January 31,2011. "Making Home Affordable Modification Program." Retrieved from http://www.treasury.gov

U.S. Treasury Department. January 2013. Making Home Affordable Modification Program." Retrieved from http://www.treasury.gov

U.S. Treasury Department. 2016. “Making Home Affordable.” *Program Performance Report through the Fourth Quarter of 2016*.
https://www.treasury.gov/initiatives/financial-stability/reports/Documents/MHA%20Quarterly%20Report%20Q4%202016_C.pdf

Warren, Elizabeth. *Taking Stock: What Has the Troubled Asset Relief Program Achieved?* DIANE Publishing, 2010

Location Factors for Mortality Rates Among African-Americans

Shahdad Naghshpour, Alabama A&M University*

Sediq Sameem, Alabama A&M University**

ABSTRACT

This study determines convergence of African-American mortality rates in the United States with a particular focus on locational differences. Using county level data during 1968-2015, the convergence rate is higher in the South among regions. While splitting the sample between the Black Belt and non-Black Belt, the higher speed of convergence is obtained in the former, and when splitting the sample between the urban and rural areas, the higher speed of convergence is obtained in the latter.

1. INTRODUCTION

The mortality rates for similar populations should be similar. The more homogenous are the populations the more similar are their mortality rates. Many of the factors that contribute to mortality can be represented by proxy variables such as location. Location is an important factor because it determines or at least represents differences in education and income, not to mention the subtler contributing factors such as socioeconomic and cultural variables. The belief is that as education spreads and medical technology becomes more universally accessible and acceptable/affordable the mortality rates should converge. There is an evidence of inherent differences in mortality rates among areas with diverse socioeconomic characteristics; for example, urban *versus* rural areas (Sameem and Sylwester 2017).

The present study divides United States into four regions of Northeast, Midwest, South, and West, following the United States Census Bureau, and compares the mortality among African-Americans in those regions for equality of the means using county level data. Convergence of mortality rates among the regions would indicate that the underlying causes of mortality rate are becoming indistinguishable among the regions. Preston (1975) provides an empirical association between per capita income and life expectancy. The Preston curve indicates that as per capita income increases life expectancy of a country increases. An increase in life expectancy means that mortality rate declines at younger age. Since everyone dies eventually, the increase in life expectancy shifts the mortality rates of younger age to older age groups; eventually, the entire population of any cohort dies.

*Department of Accounting & Finance, Alabama A&M University, Normal, AL 35762

**Department of Accounting & Finance, Alabama A&M University, Normal, AL 35762

The authors would like to thank anonymous reviewers for extensive and helpful comments and insight. All remaining shortcomings are the responsibilities of the authors.

Life expectancy increases as income increases because at higher incomes a nation can provide better health care. The association is a macro-level one and when aggregate data are decomposed to finer granules it could be that certain segments of the population of a country have higher mortality rates at a given age group than others. In other words, the conditional mortality rates could be higher for some units of population than the others.

This study explores whether mortality rates for African-Americans are converging over a period of nearly 5 decades. Using the United States county level data from 1968 through 2015 the evidence indicates β-convergence for the death rates of African-Americans. Based on various locations, convergence rate is the highest in the South, in the Black Belt region, and in rural areas.

1.1. LITERATURE REVIEW

The existence of an inverse relationship between income and mortality rate at national level does not preclude the possibility of having multitude of mortality rates at subnational levels or among the components of the national population such as ethnicity and gender. One obvious example is the differences in mortality rates among different age groups. Incomes are not the same for different regions of a country either, leading to the possibility of different mortality rates for different regions. Cutler, Deaton, and Lleras-Muney (2006) claim that the association between income and average health is stronger for poorer countries. Rogot, Sorlie, Johnson, and Schmitt (1992) demonstrate that compared to the top 5 percent of income, the life-expectancy of the people at the bottom 5 percent of income in the United States is 25 percent lower. The implication is that income convergence would contribute to the convergence of mortality rates while income divergence would hinder it.

There is a link between business cycle and mortality rate (Sameem and Sylwester 2017), which could indicate a link between economics and mortality rate. The conclusion is based on county level data from the United States. Evidence of an association between business cycle and mortality abounds; Ruhm (2000, 2015) for the United States, Neumayer (2004) for Germany, Tapia Granados (2005) for Spain, Gonzalez and Quast (2010) for Mexico, Ariizumi and Schirle (2012) for Canada, Lin (2009) for Pacific Asian countries, and Gerdtham and Ruhm (2006) for OECD countries. Other studies claim an indirect relationship between economic activities and mortality, such as pollution caused by greater economic activities (Davis *et al.* 2010). There is also evidence to differentiate pollution-caused mortality rates between rural and urban areas (Chay and Greenstone 2003, Currie and Neidell 2005, Currie and Schmieder 2009, Foster *et al.* 2009, Heutel and Ruhm 2016, Calderon-Garciduenas *et al.* 2015, and Zhou *et al.* 2015). Other studies that are conducted at a micro level demonstrate a positive association between unemployment and mortality. See Winkleman and Winkleman (1998), Burgard *et al.* (2007), Sullivan and Wachter (2009), Strully (2009) and Tapia Granados *et al.* (2014).

As demonstrated, there are numerous causes for mortality and hence difference in mortality. The present study uses theoretically identified variables that encompass and incorporate the above-mentioned causes to determine convergence.

1.2. CONVERGENCE

The research on (income) convergence starts with Kuznets (1955) who finds an association between economic growth and income inequality; as an economy grows income inequality increases at first then decreases. The concept of β-convergence begins with Baumol (1986). The common procedure for conducting β-convergence is to regress the last observation of the variable of interest on its first observation for a group of regions. Although the data consists of many observations over time for a reasonably large group of regions, the analysis is practically cross sectional instead of a panel analysis. Alternative procedures utilize other dependent variables such as the ratio of the last to the first observations, the rate of change over the period, or growth rate over time (Li and Wang 2016). Logarithmic transformations of the data are also prevalent (Parsley and Wei 1996, and Quah 1993).

There are two main streams of research in β-convergence literature: (1) unconditional convergence and (2) conditional convergence. To account for variables that are customarily assumed to be constant for the sake of theoretical simplicity control variables could be added to the analysis, which customarily is referred to as the conditional β-convergence. Although there are certain factors that influence mortality rate such as vaccinations and nutrition (Szreter 1988, Guha 1994, Fogel 1997, and Fogel 2004) or exercise (Ruhm 2000), to warrant the use of conditional convergence, lack of necessary data at the county level prohibits incorporating those factors. Other control variables include income, education, population, security measures, etc. (Barro 1991). Therefore, the analysis here is based on the use of unconditional convergence.

This study examines mortality rate convergence based on three geographical locations: (a) the United States Census Bureau regions (Northeast, Midwest, South, West), (b) Black Belt *versus* non-Black Belt, and (c) urban *versus* rural areas. There is ample evidence of differences among the regions of the United States (Nissan and Naghshpour 2014). Though most of it is part of the South, the choice of the Black Belt is for its significant historical importance. According to Britannica, the "physical region in Alabama and Mississippi, US, so named for its soil. The Black Belt is a fertile plain…A region of dark, calcareous soils, it was one of the South's most important agricultural areas before the American Civil War…cotton was the most important crop… Though strictly the name of a physical region, the term Black Belt has been borrowed by social scientists to denote those areas of the South where the plantation system, with its large number of black slaves, predominated before the Civil War." The Black Belt region includes the states of Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia. The region was the main center of slavery. Lack of convergence between the counties in the Black Belt and the other counties would indicate a prolonged and lasting consequence of slavery. If found, the

convergence would show that finally the region is being assimilated into the rest of the country, with regard to mortality rates. The authors were not able to identify a similar study about the region. Urban areas are hubs of large industries, corporations, organizations, etc., therefore, most of the labor force fluctuations occur in urban areas, increasing the likelihood of adverse health-effects as compared to rural areas. In addition, the higher level of air pollution (Calderon-Garciduenas *et al.* 2015, Zhou *et al.* 2015) and larger number of vehicles (French and Gumus 2014) in larger cities could further add to the mortalities. The negative contributions of the above factors to mortality rates of urban areas might be offset, at least in part, by better access to health care and higher income and sanitation standards. These factors might make a significant difference in mortality rates of the people living in urban versus those in rural areas.

The remainder of the paper is organized as follows: Section 2 explains data. Section 3 presents empirical methodology. Section 4 explains results, and Section 5 concludes.

2. DATA

Data are annual observations of African-American mortality rates per 100,000 population in the United States counties over the period 1968-2015, obtained from the Compact Mortality Files (CMF) of the Centers for Disease Control and Prevention (Data link: http://wonder.cdc.gov/mortsql.html). The CMF has detailed information about the people who died in the United States, classified by gender, age, race, and the cause of death. Of the over 3,000 counties in the United States, only 1,175 counties have sufficiently large enough African-American population to assure confidentiality of the deceased as well as reliability of the data. Data from all the counties of the U.S. has been utilized here but non-availability of sufficient African-American population to warrant reliable estimates brings down the size of the sample. This is why our sample of study has fewer counties. To ease comparisons, total number of observations and the total number of counties are reported for each regression. Table 1 shows average mortality rates for African-Americans in the United States.

Table 1: Average African-American Mortality Rates in the United States

	United States	Northeast	Midwest	South	West
All	953.7	736.4	853.7	1010.4	633.0
Males	1066.0	827.4	924.8	1132.8	705.9
Females	844.9	657.0	725.5	897.0	557.7
	United States	Black Belt	Non-Black Belt	Urban	Rural
All	953.7	999.9	848.6	842.7	1104.9
Males	1066.0	1127.1	915.5	949.4	1247.6
Females	844.9	894.1	721.3	748.4	998.5

Note: The mortality rates are the number of deaths per 100,000 people.

The upper panel of Table 1 displays means across the four major regions (see Appendix for details) and the United States as a whole. It is easy to note the larger share of African-American mortality rate in the South as that is where the Black Belt is located. The lower panel of Table 1 compares Black Belt *versus* non-Black Belt regions as well as urban *versus* rural areas. The United States Office of Management and Budget designates a county with a population over 50,000 as a metropolitan or urban county. It is noteworthy that the Black Belt is mostly rural, which might augment the mortality rate. Given differences in these distributions, it is plausible that other characteristics could also contribute to the difference among these areas, including the convergence rates.

3. METHODOLOGY

The model for detecting β-convergence among the mortality rates of African-Americans in the United States with a focus on locational differences is depicted in equation (1), which is a modification of the common model used in neoclassical growth literature and the law of one price (Parsley and Wei 1996, and Goldberg and Verboven 2005).

$$\Delta M_{it} = \alpha_i + \delta_t + \beta M_{it-1} + \sum_{i=1}^{s(k)} \gamma \Delta M_{it-1} + \epsilon_{it} \quad (1)$$

The model in (1) relates changes in the natural logarithm of African-American mortality rates in county *i* over sample period *t* considered, or simply the growth rate (ΔM_{it}), to the natural logarithm of the previous level of mortality rates in county *i* (M_{it-1}) and lag values of the dependent variable (ΔM_{it-1}) along with county fixed effects (α_i), time fixed effects (δ_t), and an error term (ϵ_{it}). M_{it} is defined as the percentage difference in the rate of mortality at time *t* between county *i* (AAM_{it}) and the cross-sectional average ($\overline{AAM_t}$). Mathematically, $M_{it} = \ln\left(\frac{AAM_{it}}{\overline{AAM_t}}\right)$ where AAM_{it} is the African American mortality rate in county *i* and $\overline{AAM_t}$ is the cross sectional average mortality rate, and the subscript *t* captures time. Δ is the first difference operator and $s(k)$ represents the number of lags included in our model to control for serial correlation. Akaike Information Criterion (AIC) is used to determine the number of lags. We account for potential spatial heterogeneities and seasonal effects by incorporating cross-sectional (α_i) and time (δ_i) dummies into our specifications. β is the main coefficient of interest. An ordinary least squares model is used to estimate the parameters of the model. A negative and statistically significant β with a value between zero and one means the presence of unconditional β-convergence. In order to control for possible geographical correlation between mortality rates, all specifications are estimated using clustered standard errors at the county level (Enamorado, López-Calva, and Rodríguez-Castelán 2014). When dealing with panel data, model errors in different time periods for a given cluster (county here) may be correlated, while model errors for different clusters are assumed to be uncorrelated, and failure to control for within-cluster error correlation can lead to misleading small standard errors, large t-statistics, and consequently misleading inferences (Cameron and Miller 2015).

The model in (1) has been estimated for the entire population of African-Americans in the United States as well as for the four major regions of the United States – Northeast, Midwest, South, and West, separately. We also test for differences in convergence of mortality rates between the Black Belt region *versus* non-Black Belt region. In addition, the convergence rates of urban *versus* rural counties are estimated. It is important to realize that the mortality rates could increase or decrease regardless of whether the mortality rates are converging or diverging. Therefore, a simple indication of convergence or divergence does not necessarily mean a "good" or a "bad" thing. The data need to be carefully examined to decipher trends and patterns in addition to the issues of convergence or divergence.

Furthermore, it is necessary to test for the equality of the means of mortality rates for the regions that are created using different grouping in order to determine whether there are differences in the mortality rates at the beginning of the period. It is also prudent to perform the same test for the ending date data. Once it is established that the base year mortality rates are different it makes sense to test for convergence of mortality rates. Naghshpour (2008) demonstrates that the choice of the beginning and ending dates affects the inference regarding convergence. Therefore, it would be advisable to use more than one beginning and ending combinations for determining convergence. Eff (1999) and Naghshpour (2008) use all the possible combinations of beginning and ending dates at their disposal. Here, however, a more conservative undertaking would be acceptable because the objective is to ascertain the existence of convergence of mortality rates and the determining factors explaining mortality rates.

4. RESULTS

In order to justify testing for convergence it is necessary to determine whether there was a difference in mortality rates of African-Americans at the beginning of the study period. In order to calculate the necessary rates, two data pointes are lost, therefore, the first available data is for 1970. A One-way Analysis of Variance (ANOVA) is conducted to determine the equality of the means among the regions of the country. Table 2 provides the means, standard deviations, and the frequencies of the available data for each of the regions. There is a substantial variation in the frequencies for the regions ranging from 40 (region 4) to 816 (region 3). The total number of available data is 1,051, which is far less than the number of counties in the country. One reason for elimination, as stated earlier, is lack of sufficiently large enough African-American population in a county to allow anonymous reporting. This figure is lower than the number of observations reported earlier due to the fact that the data that was indicated as "unreliable" by CDC have been deleted as well. The mean mortality rates range from the low of 821 (region 4) to 1187 (region 3).

Table 2: Test of Equality of the Mean Mortality Rates for 1970

Regions	Mean	Std. Dev.	Frequency
1 (Northeast)	964.2	317.2	82
2 (Midwest)	1130.2	496.8	113
3 (South)	1186.9	303.1	816
4 (West)	821.3	314.3	40
Total	1149.5	342.0	1,051
F	25.33		
Prob > F	0.000		

Note: F test is for the equality of the means across regions.

Table 2 provides the one-way test of equality of the means. The low *p*-value provides justification for rejecting the null hypothesis that all the means are the same in favor of the alternative that at least one mean is different than the other means.

Since the equality of the means for the regions is rejected it is necessary to determine which mean(s) is (are) different using multiple comparison tests. There are several alternative procedures for multiple comparison tests available. Table 3 displays the results for Scheffe test; the results are comparable to those of Bonferroni and Sidak (not reported here).

Table 3: Multiple Comparison Test of Equality of the Mean Mortality Rates for 1970

Mean	1 (Northeast)	2 (Midwest)	3 (South)
2 (Midwest)	166.0 (0.008)		
3 (South)	222.8 (0.000)	56.7 (0.404)	
4 (West)	-142.9 (0.171)	-308.9 (0.000)	-365.6 (0.000)

Note: Results are based on Scheffe test, but are comparable with Bonferroni and Sidak. P-values are in parentheses.

The results indicate that the mean mortality rates for regions 1 and 2 and regions 1 and 3 are different. The means for regions 2 and 4 as well as for regions 3 and 4 are also different. However, the means for regions 3 and 4 and the means for regions 2 and 3 are not statistically significantly different. In Table 4 a line under a group of regions indicates the means cannot be distinguished statistically.

Table 4: Comparisons of the Means for 1970

Region	1 (Northeast)	4 (West)	2 (Midwest)	3 (South)
Means	821	964	1130	1187

Clearly, there is a geographical discrepancy among the mortality rates. The mortality rates for regions 1 (Northeast) and 4 (West) are statistically indistinguishable; as are regions 2 (Midwest) and 3 (South). Northeast and West not only have lower rates but they also have relatively fewer African-Americans.

Table 5 presents the results of test of equality of the means for mortality rates among African-Americans by regions for 2015. The means range from 526 (region 4) to 908 (region 3). All regions have lower mortality rates in 2015 as compared to 1970. Region 2 (Midwest) has the largest decline of 459, while region 3 (South) has the lowest, equaling to 279. The total number of available data has increased from 1,051 counties to 1,127. The bottom portion of Table 5 indicates that the means of mortality rates among the regions are still different.

Table 5: Test of Equality of the Mean Mortality Rates for 2015

Regions	Mean	Std. Dev.	Frequency
1 (Northeast)	602.9	206.7	93
2 (Midwest)	671.7	292.7	139
3 (South)	907.5	273.0	829
4 (West)	526.1	195.6	66
Total	831.0	297.0	1,127
F	90.64		
Prob > F	0.000		

Note: F test is for the equality of the means across regions.

Table 6 presents the results of comparison tests. Again the results for Scheffe test are presented. The means for regions 1 (Northeast) and 2 (Midwest) are no longer differentiable statistically. As before, the means for regions 1 (Northeast) and 4 (West) are not distinguishable at conventional statistical significance levels. The remaining differences are all significant, statistically.

Table 6: Multiple Comparison Test of Equality of the Mean Mortality Rates for 2015

Mean	1 (Northeast)	2 (Midwest)	3 (South)
2 (Midwest)	68.8 (0.296)		
3 (South)	304.6 (0.000)	235.9 (0.000)	
4 (West)	-76.8 (0.362)	-145.6 (0.004)	-381.5 (0.000)

Note: Results are based on Scheffe test, but are comparable with Bonferroni and Sidak. P-values are in parentheses.

The summary of the means are as follows; a line under regions indicates equality of the corresponding means.

Table 7: Comparisons of the Means for 2015

Regions	4 (West)	1(Northeast)	2 (Midwest)	3 (South)
Means	525	603	672	908

Comparing Tables 4 and 7 indicates that the reduction of the mortality rates between 1970 and 2015 in region 2 (Midwest) was substantially greater than that of region 3 (South) to make the means different. Therefore, although all mortality rates declined, the South was not as successful as the Midwest, where its mortality rate is no longer distinguishable from that of Northeast.

Mortality rates have declines substantially in all regions. The declines are, from highest to the lowest, 40.5%, 37.5%, 36.1%, and 23.5% for the Midwest, Northeast, West, and South, respectively. The decline for Northeast and West are very close, while the decline in the mortality rate in the Midwest is more substantial. The decline in the South, although considerable, is much less than the rest of the regions. Consequently, the distinction between the mortality rates of the Midwest and Northeast have vanished, although Midwest rate is still statistically significantly different from that of the West; but the gap is closing. The fact that the mortality rate of the South is declining much slower than the ones for the other regions could result in divergence of the mortality rates among the regions. However, the slower rate of decline in the South is not sufficient to cause lack of convergence in mortality among the regions, let alone to produce a statistically significant divergence, yet.

Next, we switch to the results obtained from model (1) for β-convergence. The upper panel of Table 8 reports baseline results from model (1) when considering the African-American mortality rates for the United States as a whole (column 1) and for different regions (columns 2-5) during the sample period, 1970-2015. The estimated β coefficients are negative, between zero and one, and highly statistically significant in all specifications considered. They are consistent with the β-convergence requirement.

Among the regions, the greatest degree of convergence is in the South. The fact that the mortality rates are converging in all regions indicates that contributing factors are not limited to one region or another. The exact contributing factor at the national level are beyond the scope of the present study.

The middle and the lower panels of Table 8 report results for males and females, respectively. The estimated coefficients are negative indicating β-convergence for each group. However, the coefficients for females are much larger in magnitude in comparison to those for males, suggesting a faster rate of convergence among female mortality rates. Based on gender, the speed of convergence is the highest for males in the Midwest, and for females in the South. Since a population consist of different components, such as male and female that are known to have distinct mortality rates, it would be prudent to examine whether the mortality rates for each component is converging as well. The mortality rates for both male and

female African-Americans are converging in all regions as well as the entire country. Therefore, there is no gender bias in mortality convergences.

Table 9 extends the analysis to comparisons between the Black Belt region and non-Black Belt region (columns 2-3) as well as between urban and rural areas (columns 4-5). For ease of the comparisons, the results for the United States as a whole are shown in column 1. The speed of convergence, as measured by the magnitude of β-coefficient, is higher for African-American mortality rates in the Black Belt region and the rural areas.

Table 8: β-Convergence of African-American Mortality Rates in United States Regions

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	United States	Northeast	Midwest	South	West
All	-0.392***	-0.324***	-0.417***	-0.419***	-0.203***
	(0.014)	(0.035)	(0.035)	(0.017)	(0.025)
# Observations	45,707	3,605	4,840	35,179	2,083
# Counties	1,175	91	145	875	64
Males	-0.424***	-0.384***	-0.485***	-0.443***	-0.246***
	(0.017)	(0.034)	(0.045)	(0.021)	(0.034)
# Observations	35,071	3,094	3,575	26,805	1,597
# Counties	952	78	102	725	47
Females	-0.505***	-0.486***	-0.474***	-0.525***	-0.354***
	(0.017)	(0.045)	(0.040)	(0.020)	(0.053)
# Observations	31,897	2,924	3,239	24,433	1,301
# Counties	890	76	94	680	40

Notes: Average rate of mortality for each subgroup is used as the benchmark. All regressions contain county and year fixed effects. Sample period is 1970-2015. Clustered standard errors at the county level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: β-Convergence of African-American Death Rates in Black Belt & Urban-Rural Areas

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	United States	Black Belt	Non-Black Belt	Urban	Rural
All	-0.392***	-0.419***	-0.340***	-0.309***	-0.582***
	(0.014)	(0.019)	(0.019)	(0.014)	(0.016)
# Observations	45,707	32,304	13,403	26,738	18,969
# Counties	1,175	788	387	678	497
Males	-0.424***	-0.452***	-0.356***	-0.345***	-0.614***
	(0.017)	(0.023)	(0.020)	(0.018)	(0.025)
# Observations	35,071	24,943	10,128	21,878	13,193
# Counties	952	667	285	572	380
Females	-0.505***	-0.533***	-0.435***	-0.421***	-0.709***
	(0.017)	(0.022)	(0.025)	(0.019)	(0.019)
# Observations	31,897	22,717	9,180	20,137	11,760
# Counties	890	628	262	544	346

Notes: See notes to Table 8. Black Belt includes Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia. Urban (Rural) counties are those with more than (less than) 50,000 people. *** p<0.01, ** p<0.05, * p<0.1

5. CONCLUSION

Applying the analysis of convergence to the trends of mortality rates, this study finds the presence of β-convergence among the African-American population in the United States. Using county level annual mortality rates data during 1970-2015, the results suggest that mortality rates of African-Americans are mean-reverting. In other words, they are moving towards the cross-sectional average. The results are stronger for females, particularly in rural areas and the Black Belt region. It is not clear, however, what is the cause of convergence of mortality rates in different regions. Mortality rates could converge when the mortality rates decrease for regions with high mortality rates. The cause, however, could be due increases in the mortality rates of areas with low mortality.

One of the limitations of the study is lack of control variables that may have significant impact on mortality rate at the county level. Second, the use of county level data itself is prone to errors as Pierce and Denison (2006) show reporting errors from Texas using county level data whereas larger units of analysis such as those at state level data are likely to better filter out random errors. At the same time, the advantage of using county level data is that within county variation is likely to be smaller than within state variation thus allowing for less heterogeneity within the unit of analysis. In addition, the larger number of observations at the county level can increase the power of statistical tests.

REFERENCES

Ariizumi, H. and Schirle, T. 2012. "Are recessions really good for your health? Evidence from Canada." *Social Science & Medicine*, 74(8): 1224-1231.

Barreto, R.A. and Hughes, A.W. 2004. "Under performers and over achievers: a quantile regression analysis of growth." *Econ Record*, 80(248), 17–35.

Barro, R.J. 1991. "Economic growth in a cross section of countries." *The Quarterly Journal of Economics*, 106(2), 407-443.

Barro, R. and Sala-i-Martin, X. 1991. "Convergence Across Sates and Regions." *Brooking Papers on Economic Activity*, 107-158.

Britannica. https://www.britannica.com/place/Black-Belt. Accessed July 14, 2018.

Burgard, S.A., Brand, J. E., and House, J. S. 2007. "Toward a better estimation of the effect of job loss on health." *Journal of Health and Social Behavior*, 48(4): 369-384.

Calderon-Garciduenas, L., Kulesza, R.D.R., D'Angiulli, A., and Torres-Jardon, R. 2015. "Megacities Air Pollution Problems: Mexico City Metropolitan Area Critical Issues on the Central Nervous System Pediatric Impact." *Environmental Research*, 137: 157-169.

Cameron, A.C. and Miller, D. L. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317-372.

Chay, K. and Greenstone, M. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics*, 118(3): 1121-1167.

Clutter, D., Deaton A., and LLeras-Muney, A. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives*. 20(3), 97-120.

Cook, S. and Winfield, T. 2012. "Crime across the States: Are US crime rates converging?" *Urban Studies*, 0042098012466602.

Currie, J. and Schmieder, J. 2009. "Fetal Exposure to Toxic Releases and Infant Health." *American Economic Review*, 99(2): 177-183.

Currie, J. and Neidell, M. 2005. "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *Quarterly Journal of Economics*, 120(3): 1003-1030.

Davis, M.E., Laden, F., Hart, J.E., Garshick, E., and Smith, T.J. 2010. "Economic activity and trends in ambient air pollution." *Environmental Health Perspectives*, 118(5): 614.

Eff, E.A. 1999. "Myrdal Contra Ohlin: Accounting for the Sources of U.S. County Per Capita Income Convergence Using a Flexible Decomposition Approach." *Review of Regional Studies*, 29: 13-36.

Enamorado, T., López-Calva, L.F., and Rodríguez-Castelán, C. 2014. "Crime and growth convergence: Evidence from Mexico." *Economics Letters*, 125(1): 9-13.

Farbman M. 1973. "Income Concentration in the Southern United States." *Review of Economics and Statistics,* 55(3): 333-340.

Fogel, R.W. 2004. "The Escape from Hunger and Premature Death. 1700-2100." Cambridge: Cambridge University Press.

Fogel, R.W. 1997. "New Findings on Secular Trends in Nutrition and Mortality: Some Implications for Population Theory." *In Handbook of Population and Family Economics*. Mark R. Rosenzweig and Oded Stark, eds. New York: Elsevier Science, North Holland. 433-481.

Foster, A., Gutierrez, E., and Kumar, N. 2009. "Voluntary Compliance, Pollution Levels, and Infant Mortality in Mexico." *American Economic Review*, 99(2): 191-197.

French, M. and Gumus, G. 2014. "Macroeconomic Fluctuations and Motorcycle Fatalities in the U.S." *Social Science & Medicine*, 104: 187-193.

Gerdtham, U.G. and Ruhm, C.J. 2006. "Deaths rise in good economic times: evidence from the OECD." *Economics & Human Biology*, 4(3): 298-316.

Goldberg, P.K. and Verboven, F. 2005. "Market integration and convergence to the Law of One Price: evidence from the European car market." *Journal of International Economics*, 65(1): 49-73.

Gonzalez, F. and Quast, T. 2010. "Mortality and business cycles by level of development: Evidence from Mexico." *Social Science & Medicine*, 71(12): 2066-2073.

Granados, J.A.T. 2005. “Recessions and mortality in Spain, 1980–1997.” *European Journal of Population*, 21(4): 393-422.

Guha, S. 1994. “The Importance of Social Intervention in England’s Mortality Decline: The Evidence Reviewed.” *Social History of Medicine*. 7(1): 89-113.

Heutel, G. and Ruhm, C.J. 2016. “Air pollution and procyclical mortality.” *Journal of the Association of Environmental and Resource Economists*, 3(3): 667-706.

Kuznets, S. 1955. “Economic growth and income inequality.” *The American Economic Review*, 45(1): 1-28.

Li, X. and Wang, R. 2016. “Are US obesity rates converging?” *Applied Economics Letters*, 23(8): 539-543.

Lin, S.J. 2009. “Economic fluctuations and health outcome: a panel analysis of Asia-Pacific countries.” *Applied Economics*, 41(4): 519-530.

Murray, Barbara, B. 1970. “Coefficient of Intra area Variation as a Measure of Spatial Income Inequality.” *Journal of the American Statistical Association*, 65: 598-601.

Naghshpour, S. 2008. “The Influence of Temporal Choice, Correlation, and Sample Size on Income Convergence.” *Review of Regional Studies*, 38(2): 211-231.

National Center for Health Statistics, Compressed Mortality File (CMF 1999-2013, Series 20, No. 2S, 2014, CMF 1989-1998, Series 20, No. 2E, 2003, CMF 1968-1988, Series 20, No. 2A, 2000), as compiled from data provided by the 57 vital statistics jurisdiction through the Vital Statistics Cooperative Program.

Neumayer, E. 2004. “Recessions lower (some) mortality rates: evidence from Germany.” *Social Science & Medicine*, 58(6): 1037-1047.

Parsley, D.C. and Wei, S.J. 1996. “Convergence to the Law of One Price without trade barriers or currency fluctuations.” *The Quarterly Journal of Economics*, 111(4): 1211-1236.

Pierce, R.J. and Denison, A.V. 2006. “Place-of-residence errors on death certificates for two contiguous US counties.” *Population Health Metrics*, 4(1): 6.

Preston, S.H. 1975. “The changing relation between mortality and level of economic development.” *Population Studies*, 29(2): 231-248.

Quah, D. 1993. “Galton's Fallacy and Tests of the Convergence Hypothesis.” *Scandinavian Journal of Economics*, 95: 427-443.

Rogot, E., Sorlie, P.D., Johnson, N.J., and Schmitt. C. 1992. “A Mortality Study of 1.3 Million Persons.” Bethesda, MD: National Institutes of Health, National Heart, Lung, and Blood Institute.

Ruhm, C.J. 2015. “Recessions, healthy no more?” *Journal of Health Economics*, 42: 17-28.

Ruhm, C.J. 2000. “Are Recessions Good for Your Health?” *The Quarterly Journal of Economics*, 115(2): 617-650.

Sameem, S. and Sylwester, K. 2017. “The business cycle and mortality: Urban versus rural counties.” *Social Science & Medicine*, 175: 28-35.

Strully, K.W. 2009. “Job loss and health in the US labor market.” *Demography*, 46(2): 221-246.

Sullivan, D. and Von Wachter, T. 2009. “Job displacement and mortality: An analysis using administrative data.” *The Quarterly Journal of Economics*, 124(3): 1265-1306.

Szreter, S. 1988. “The Importance of Social Intervention in Britain’s Mortality Decline 1859-1940.” *Social History of Medicine*. 1(1): 1-37.

Tapia Granados, J.A., House, J.S., Ionides, E.L., Burgard, S., and Schoeni, R.S. 2014. “Individual joblessness, contextual unemployment, and mortality risk.” *American Journal of Epidemiology*, 180(3): 280-287.

Winkelmann, L. and Winkelmann, R. 1998. “Why are the unemployed so unhappy? Evidence from panel data.” *Economica*, 65(257): 1-15.

Zhou, M., He, G., Liu, Y., Yin, P., Li, Y., Kan, H., Fan, M., Xue, A., and Fan, M. 2015. “The Associations between Ambient Air Pollution and Adult Respiratory Mortality in 32 Major Chinese Cities, 2006-2010.” *Environmental Research*, 137: 278-286.

APPENDIX

Table 1A: US States by Regions

Northeast	Midwest	South	West
Division I	**Division III**	**Division V**	**Division VIII**
Connecticut	Indiana	Delaware	Arizona
Maine	Illinois	District of Columbia	Colorado
Massachusetts	Michigan	Florida	Idaho
New Hampshire	Ohio	Georgia	New Mexico
Rhode Island	Wisconsin	Maryland	Montana
Vermont		North Carolina	Utah
	Division IV	South Carolina	Nevada
Division II	Iowa	Virginia	Wyoming
New Jersey	Kansas	West Virginia	
New York	Minnesota		**Division IX**
Pennsylvania	Missouri	**Division VI**	Alaska
	Nebraska	Alabama	California
	North Dakota	Kentucky	Hawaii
	South Dakota	Mississippi	Oregon
		Tennessee	Washington
		Division VII	
		Arkansas	
		Louisiana	
		Oklahoma	
		Texas	

Source: US Census Bureau: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Table 2A: β-Convergence of African-American Mortality Rate during Post-Great Recession

	(1)	(2)	(3)	(4)	(5)
	United States	Northeast	Midwest	South	West
All	-0.022***	0.023***	-0.212***	-0.061***	-1.056***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.191)
# Observations	8,203	675	937	6,157	434
R Squared	0.033	0.043	0.041	0.033	0.039
Males	-0.025***	-0.007***	-0.296***	-0.361***	-0.065***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
# Observations	6,245	594	688	4,640	323
R Squared	0.039	0.056	0.037	0.041	0.063
Females	-0.025***	0.038***	-0.170***	0.023***	-0.008
	(0.000)	(0.000)	(0.000)	(0.000)	(0.052)
# Observations	5,934	567	602	4,470	295
R Squared	0.034	0.034	0.041	0.035	0.079

Notes: Average rate of mortality for each subgroup is used as the benchmark. All regressions contain county and year fixed effects. Sample period is 2009-2015. Clustered standard errors at the county level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The Minutes of the FOMC: How Economic Factors Influence the Language of Federal Reserve Chairs

Scott Miller[1]

Victor Philaire**[2]

ABSTRACT

This paper uses the Linguistic Inquiry and Word Count (LIWC) program to perform a textual analysis of the Federal Open Market Committee (FOMC) Minutes. The main objective of this research is to examine the impact of various economic factors on the amount and type of language used by the Federal Reserve. We compare the results over three different Fed chairs (Greenspan, Bernanke and Yellen) from 1993 to 2017. Our findings suggest that each of the Federal Reserve Chairs portray the state of the economy in an empirically distinctive manner.

INTRODUCTION

The U.S. Federal Reserve is arguably one of the most powerful economic institutions around the globe, which is why the economic world is so closely attentive to the smallest details or pieces of information that it provides.

Markets, however, can respond asymmetrically to new announcements. Bad news will have a deeper impact on markets than the publication of positive news and forecasts (Anderson et al., 2003). Hence, while their words were powerful, Federal Reserve Chair Alan Greenspan and his successors needed to carefully weigh the words they employed. The Federal Open Market Committee (FOMC) is the institution within the Federal Reserve responsible for the nation's open market operation. Ever since the recession of 2008 and near zero interest rates, the Federal Reserve and the FOMC have lost their ability to apply monetary accommodations on the lower-end of the interest rate spectrum. Stripped of this ability, the forward guidance from the public statements of the FOMC (Boukus and Rosenberg, 2006) serves as a substitute to influence the economy and the markets on a short-term macroeconomic scale (Campbell et al., 2012).

Ever since the end of the twentieth century, communication has become one of the most important aspects of monetary policy. The central banks, and more specifically the Federal Reserve, increasingly witness the growing importance of the content and style of their communications. However, the strong

[1] University of Tampa, Sykes College of Business, 401 W Kennedy Blvd, Tampa, FL 33606. samiller@ut.edu

[2] University of Tampa, Sykes College of Business, 401 W Kennedy Blvd, Tampa, FL 33606. victor.philaire@spartans.ut.edu

variation in communication strategies between the central banks across the globe (Blinder et al., 2008) creates the question of whether this situation is solely central bank specific. For the U.S. Federal Reserve, there could be a difference in communication strategy simply because of the person presiding on the board. However, this could also be the result of macroeconomic factors.

In any case, having a deeper understanding of the communication style of individual Federal Reserve Chairs is important. It can assist all market participants in better interpreting the underlying meaning of the messages being sent. This study will provide valuable insight in that regard.

LITERATURE REVIEW

This paper is related to three different threads of existing literature. The first deals with linguistic analysis. Tausczik and Pennebaker (2010) examine the growing field of computerized text analysis methods and argues for the reliability and transparency of the Linguistic Inquiry and Word Count (LIWC) program at referencing terms and strings while also providing useful empirical evidence for researchers. The study was employed to reveal evidence of the influence that existed within the FOMC forecasts of various economic indicators and the policy implications of the signals provided by these forecasts.

The second thread relates to the significant impact an institution's communication can have at a macroeconomic scale. Anderson et al. (2003) examine the asymmetric impact of positive and negative news on the macroeconomy. They provide empirical evidence that large institutions such as the government are powerful enough to put the markets, or the economic environment as a whole, in significant turmoil after publishing negative information. However, positive news will generate a less substantial reaction on the up-side. Blinder et al. (2008), Boukus and Rosenberg (2006), and Campbell et al. (2012) study the implications of central bank communication strategies with a more forward-looking approach. They provide significant evidence that central banks' communication can have a non-negligible impact on economic factors, including changes in inflation, and spikes in economic uncertainty. Sicilia and Cruikshank (2000) analyze how the words and body language of the Federal Reserve chairman can have a significant impact on the markets but they also show that recurring unexpected news and surprises can amplify the attention given to the smallest details. This research was proven to be particularly accurate during the time of Alan Greenspan as the chair of the Federal Reserve, as will be discussed later.

In the third thread of the existing literature, Costa and Jacob (2011), explain and define the S&P 500 Index as a suitable benchmark to be employed when developing macroeconomic research.

DATA

Data collection

Since 1933, the Federal Reserve has published the minutes of the FOMC meetings. These meetings are held eight times each year to review the economic and financial conditions, determine the appropriate stance of monetary policy, and assess the risks to the long-run goals of price stability and sustainable economic growth. This study analyzes the minutes of the FOMC from the period of Alan Greenspan's

meeting on February 3rd, 1993 to Janet Yellen's meeting on July 26th, 2017. The minutes, ranging from the years 1993 to 2017, have been collected through a Bloomberg Terminal and from the Federal Reserve's website. The FOMC minutes in their current form emerged in 1993. From 1967 to 1992, the precursor to modern minutes were formatted as "Records of Policy Actions & Minutes of Actions". From 1993 to present, the FOMC minutes are in a consistent form and provide a summary of issues addressed at each FOMC meeting. Therefore, we start our sample period after this transition in February 1993 to preserve the integrity of the data and maintain consistency.

The market data was extracted from the Center for Research in Security Prices (CRSP) dataset. The information utilized contains data related to the S&P 500 Composite Index. For the purpose of this study, the S&P 500 is employed as the appropriate market benchmark (Costa and Jacob, 2011).

The seasonally adjusted real Gross Domestic Product (GDP) data was downloaded from the St Louis Fed's website. The real GDP growth is computed from the change in real GDP between each quarter.

The monthly seasonally adjusted Consumer Price Index (CPI) data for all urban consumers was downloaded from the St Louis Fed's website as well. The inflation rate employed for the research is computed through the use of the monthly seasonally adjusted CPI for all urban consumers to capture the change in price levels across the time period of the study.

The monthly Effective Federal Funds Rate (FFR), the monthly 3-Month Treasury Bill, the monthly 6-Month Treasury Bill, the monthly 1-Year Treasury Bill, the monthly 3-Year Treasury Note, the monthly 5-Year Treasury Note, and the monthly 10-Year Treasury Note were also downloaded through the St Louis Fed's website.

The monthly seasonally adjusted Unemployment Rate in the United States was also downloaded from the St Louis Fed's website.

Data Formatting

The Federal Reserve archives all the meetings, policies, transcripts, minutes and other documents related to the Federal Open Market Committee meetings from 1933 until 2017. All of the minutes of the FOMC meetings held between January 1993 and September 2017 were acquired for use in this study. Each of the minutes was then grouped under the three different chairs who successively held office throughout the period: A. Greenspan (01/1993-01/2006), B. Bernanke (02/2006-01/2014), and J. Yellen (02/2014-09/2017).

The minutes were studied using the Linguistic Inquiry and Word Count (LIWC) program (Tausczik and Pennebaker, 2010). The program searches for specific terms and strings defined in a comprehensive dictionary. The dictionary groups the terms and strings into multiple categories. For the purpose of this study, three categories were retained and used in our empirical analysis: word counts, positive words, and negative words. The word count provides insight on the volume of information the board of the FOMC decides to communicate in regard to the current economic environment. The positive words variable measures how markedly the board's communication language is conveying a positive tone while the

negative words variable weighs how significantly the minutes are presenting a more unfavorable message in regard to the economy. Both word outputs are percentages of the total number of words employed in the minutes of each meeting. The results from the linguistic analysis were incorporated into multiple files which followed the same grouping, based on the presiding chair at the given time, as the original downloaded records. Additionally, each of the results were matched with the dates at which the relevant FOMC meeting occurred to maintain a precise timeline throughout the research.

The economic variables employed throughout this study were downloaded individually and then merged into the same file. To ensure the timeliness of the data, the dates when the FOMC meetings were held were matched to the corresponding economic variables' value. The data collected was grouped based on the chair who was presiding over the Federal Open Market Committee meeting.

The data generated throughout the linguistic analysis of the minutes and the economic data collection were then merged, while still being grouped by the corresponding chair of the Federal Reserve. The three chairs' correlation matrices were computed through the analysis of these merged files.

Furthermore, the matrices were employed as the source file for the multiple regression analysis. The results published in Tables 3 to 6 are the five best models obtainable based on R-square computations for the 5 factors multiple regressions. The regressions identify which economic factors had the most impact on how each of the chairs and the board of the FOMC would explain the economic condition when they published the minutes at a given point in time. The resulting output is three 5 factors multiple regressions, one for each of the chairs, attempting to best predict how much they will write about the subject (word count), how optimistic (positive words), and how negative they feel about the economy.

Each of the results from the linguistic analysis is paired between two of the Federal Reserve chairs according to the following format: Greenspan/Bernanke, Greenspan/Yellen, and Bernanke/Yellen. The pairing allowed to run a t-test, two sample assuming unequal variance, to determine if the communication language employed by each of the chairs is standardized by the minutes. The t-test is employed as a control method to ensure the quality of the study.

METHODOLOGY

This paper makes two distinct contributions to the literature. First, the actual language used by each of the Fed Chairs is compared across the cross-sectional time period. The objective is to see if there are significant differences between the amount and types of language used by each Fed chair. Secondly, this paper examines how effectively this language interprets the economic conditions of the US.

Testing the difference in language across Fed chairs is done through two methods. First, descriptive statistics are run on the key language variable used. Table 1 Panel A (see Appendix) displays the descriptive statistics of the word counts and word categories used. While he presided the FOMC meetings for over thirteen years, Alan Greenspan was the most concise of the three chairs with an average of 4702.50 words per Minutes document published. He also accounts for the minimum number of words employed in a Minutes document with 2851.00 and the lowest standard deviation (1405.17). These numbers are

reflective of Greenspan's well known straightforward and terse communication. On the other hand, Janet Yellen is described as the wordiest of the three with an average of 8506.14 words employed per Minutes document. She also is the chair who used the most words in a single Minutes document and has a minimum number of words employed (6063.00) that is above the average number of words per Minutes document for all of the chairs (5871.21). Yellen has often been depicted and criticized for her extensive and verbose communication, which is corroborated by these numbers.

The second panel (B) of Table 1, displays the descriptive statistics of the positive words as a percentage of the total number of words employed in a Minutes document. While he chaired the Federal Reserve and the FOMC during the great recession of 2007, Ben Bernanke is the chair with the highest average percentage (2.49%) as well as the maximum percentage (3.22%) of positive words employed in a Minutes document. This is interesting since he chaired the Fed during what is described as the greatest financial turmoil of the twenty-first century. Conversely, Alan Greenspan, who was the chair of the Federal Reserve during the tragic terrorist attacks on US soil on September 11, 2001, accounts for the lowest percentage (1.30%) of positive words employed in a Minutes document. Janet Yellen, who has been the chair of the Fed over a consistently strengthening and growing economy, records a minimum percentage (1.89%) of positive words that is above the two other chairs but also has the lowest standard deviation (0.21).

The third panel (C) of Table 1, displays the descriptive statistics of the negative words as a percentage of the total number of words employed in a Minutes document. Following the same economic conditions aforementioned, Alan Greenspan accounts for the highest (2.76) and the lowest (0.58) percentage of negative words per Minutes document, well above and below the average across all of the chairs (1.24). He also displays the highest standard deviation (0.42), which could largely be explained by the significance of some unforeseeable events that occurred between 1993 and 2006, such as the various wars across the globe, the significantly growing economy toward the end of his mandate, etc. Once again, conversely, Ben Bernanke did not seem to excessively employ negative terms during the 2007 financial crisis since his maximum percentage of negative words (1.87) sits well below Greenspan's and is similar to Yellen's (1.59).

The second method is to run successive hypothesis tests to determine if there are significant differences in language by each of the Fed chairs. The following difference in means of unequal variances t-test hypotheses are tested for Total Word Count and categorical variables of Positive Words and Negative Words.

$$\begin{aligned} &H0: \text{Greenspan} = \text{Bernanke} \\ &H1: \text{Greenspan} \neq \text{Bernanke} \end{aligned} \qquad (1)$$

$$\begin{aligned} &H0: \text{Greenspan} = \text{Yellen} \\ &H1: \text{Greenspan} \neq \text{Yellen} \end{aligned} \qquad (2)$$

$$\begin{aligned} &H0: \text{Bernanke} = \text{Yellen} \\ &H1: \text{Bernanke} \neq \text{Yellen} \end{aligned} \qquad (3)$$

If the p-value of this hypothesis test is less than 0.10, then the null hypothesis can be rejected and conclude that there are significant differences between the two Fed chairs.

To test the second objective of the paper, two approaches are again used. The first approach uses a correlation matrix to see the relationship across all variables used in this paper. Table 2 shows the results of this matrix. The strong negative relationships (-1 > x > -0.70) and strong positive relationships (0.70 < x < 1) were both indicated using three asterisks next to the corresponding output values. For instance, all of the U.S. treasury bonds and the Federal Funds Rate have a strong positive relationship which is essentially due to the fact that the bonds rates are adjusted by the market in response to the Fed's guidance for the Federal Funds Rate. The moderate negative relationships (-0.70 > x > -0.50) and moderate positive relationships (0.50 < x < 0.70) are both identified with two asterisks next to the corresponding output values. For instance, for all Fed chairs, unemployment seemed to have a moderate negative relationship with the treasury bonds and the Federal Funds Rate. Additionally, for both Greenspan and Bernanke, the treasuries and FFR had a moderate negative relationship with the number of positive words employed as a percentage of the total number of words in a Minutes document. Ultimately, the independent variables which had weak negative relationships (-0.50 > x > -0.30) and weak positive relationships (0.30 < x < 0.50) were both indicated with one asterisk along the corresponding output values. Hence, for the period of the study, the correlation matrix indicates that real GDP growth had a weak positive relationship with the treasuries and the Federal Funds Rate. It also helps to identify a weak relationship between total number of words employed in the Minutes documents and the long-term (10Year) U.S. Treasury bond.

The second approach was to run several multivariate regressions to predict the effect economic variables have on the propensity of the Federal Reserve Chair to use certain language. The first regression run is to determine if these economic variables have an impact on the "amount" of words used by the Fed chairs. To standardize the Total Word Count Variable as a percentage, it is divided by the maximum value across the sample. The second regression tests the impact on the percent of positive emotion words and the third regression tests the impact on the percent of negative emotion words.

$$\frac{WC}{Max\ WC_t} = \beta_0 + \beta_1(GDP_{t-1}) + \beta_2(S\&P_{t-1}) + \beta_3(Inf_{t-1}) + \beta_4(UE_{t-1}) + \beta_5(IR_{t-1}) + \varepsilon_t \quad (4)$$

$$PosEmo_t = \beta_0 + \beta_1(GDP_{t-1}) + \beta_2(S\&P_{t-1}) + \beta_3(Inf_{t-1}) + \beta_4(UE_{t-1}) + \beta_5(IR_{t-1}) + \varepsilon_t \quad (5)$$

$$NegEmo_t = \beta_0 + \beta_1(GDP_{t-1}) + \beta_2(S\&P_{t-1}) + \beta_3(Inf_{t-1}) + \beta_4(UE_{t-1}) + \beta_5(IR_{t-1}) + \varepsilon_t \quad (6)$$

where:

$\frac{WC}{Max\ WC_t}$ = Percentage of Word Count to Maximum Word Count across all minutes

$PosEmo$ = Percentage of Positive Words to Word Count in each minutes

$PosEmo$ = Percentage of Negative Words to Word Count in each minutes

GDPt-1 = Real GDP Growth

S&Pt-1 = S&P 500 Growth

Inft-1 = Inflation Rate

UEt-1 = Unemployment Rate

IRt-1 = Interest Rate (FFR, T-Bill, etc.)

All independent variables are lagged for the period prior to the actual release of the Federal Reserve minutes (at time t). This allows the analysis to determine the effect the current economic conditions at the time of the meetings has on the actual amount or type of language used in the Minutes.

RESULTS

The results of this study show that major differences exist in the amount and type of language used by the different Fed Chairs. Table 1 displays the significance and details of these differences.

Further analyzing this first objective of the paper, Tables 7 through 9 show the results of Models (1) – (3), which identifies the difference in means t-test of unequal variances. The results suggest that the Federal Reserve minutes contain a substantially larger total word count when Janet Yellen was chair than when led by Ben Bernanke and Alan Greenspan. These results are both statistically significant at the <0.01 level. The results also indicate that the minutes include significantly more positive words as a percentage of the total word count per minutes when Bernanke was the chairman of the Federal Reserve than both Alan Greenspan and Janet Yellen. Again, these results are statistically significant at the <0.01 level. Finally, the results also specify a meaningful difference in the mean of negative words as a percentage of the total number of words per Minutes document when Janet Yellen was leading the Fed than when led by Greenspan (sig 0.0142) and Bernanke (sig 0.0124). These results are statistically significant at the <0.01 level. However, the analysis does not provide significant evidence to reject the null when comparing the difference in the mean for the number of negative words as a percentage of the total word count per Minutes document between Greenspan and Bernanke (sig 0.9771). For robustness purposes, we confirmed these relationships by running dummy variable regressions that control for various economic performance variables (real GDP growth, S&P 500 growth, inflation, unemployment and the federal funds rate). We find the same statistically significant relationships from the difference in means tests.

The second objective of the paper focuses on how the Federal Open Market Committee and the chair presiding over it interprets the state of the economy in its Minutes documents. The ordinary least squares (OLS) multi-variate regression models obtained for all and also for each individual chair aims to predict the percentage of maximum word count, the percentage of positive words, and the percentage of negative words employed in the Minutes documents. Each Table 3-6 uses various transformations of these regressions. Panel A displays the results of the Percentage of Total Word Count calculation (Equation 4), Panel B displays the results of the Percentage of Positive Words calculation (Equation 5) and Panel C displays the results of the Percentage of Negative Words calculation (Equation 6).

The multiple models presented in Table 3 display the independent variables employed to predict the output for each of the dependent variables with the level of statistical significance. Hence, for all chairs (Panel A of Table 3), Model 1 shows that the Federal Funds Rate is the most significant variable (sig <

0.01) and has a negative relationship with the total word count. This suggests the FOMC minutes are longer when the FFR is lowered, which seems reasonable since lowering the FFR is typically in response to some level of economic instability. However, when FFR is taken out of the equation, real GDP growth and unemployment rate become respectively a moderate negative (sig < 0.05) and strong positive (sig < 0.01) predictor of total word count. Additionally, instead of having a negative impact (Beta = -0.117) on the dependent variable (WC%), unemployment rate is positively affecting it (Beta = 3.018). This is partially explained by the negative correlation between both the unemployment rate and the FFR (Table 2 – Panel A). For all chairs (Panel B), the first model shows that real GDP growth and FFR are both significant factors of prediction (sig < 0.01) for the use positive words, and unemployment rate only becomes significant if the FFR is taken out of the regression. For the use of negative words for all chairs (Panel C), the only significant relationship is with real GDP growth. Notice that both Panels B and C show that real GDP growth has a statistically significant negative relationship with the usage of both positive *and* negative words. While it may seem unclear why there is such variability in the speech patterns when economic news is reported, it may be due to an attempt of the FOMC to stabilize markets. In response to bad economic data, the FOMC may be using more negative words to describe the data, but immediately following this with positive language to soften the market's possible backlash to the news. Alternatively, the FOMC may avoid using less descriptive words (positive or negative) upon the release of good economic data and let this data speak more for itself when presenting the information.

As detailed in Table 4, Alan Greenspan's communication is intrinsically linked to the unemployment rate and the Federal Funds rate for both the total word count and the percentage of positive words employed in a FOMC Minutes document (sig < 0.01). When the unemployment rate decreases, the word count of Greenspan decreases, but his use of positive words increase. However, there is not a significant change in the amount of negative words used. This suggests that Greenspan is more likely to use concise, optimistic language when unemployment is low and longer, less positive language when unemployment rises. The percentage of negative words in Greenspan's communication is closely related to the inflation rate (sig < 0.05) and the Federal Funds Rate (sig < 0.01). This implies that Greenspan's Fed used less negative language when inflation was rising and more negative language when inflation was falling. While this result may seem somewhat counterintuitive, it is consistent with many of Greenspan's speeches and stories in the financial press during his tenure. Greenspan viewed a deflation as a much bigger risk to economic growth than inflation.[1]

The models displayed in Table 5 show the output for each of the three defined dependent variables under Ben Bernanke as chair of the Fed. The model predicting the percentage of maximum word count indicates that both the inflation rate and the Federal Funds Rate are meaningful factors (sig < 0.1 and 0.01). When the Federal Funds rate is dropped in model (2), we find that the unemployment rate is the only significant predictor of the total word count. This suggests that the minutes were longer when unemployment rose. The results of Panels B and C are of particular interest. We find that Real GDP has

a statistically significant negative relationship with the usage of both positive and negative words. As previously mentioned, this would suggest that the FOMC is attempting to stabilize markets, particularly during the financial crisis. Bernanke's Fed may have used more negative words to describe the economic climate, but coupled these words with more optimistic, positive words to present a narrative that would calm financial markets. There is a significantly higher usage of positive words with rising unemployment, which would also support this interpretation. Bernanke's Federal Reserve in effect attempted to put a positive "spin" on negative economic news in an effort to prevent widespread panic of the worsening economic conditions at the time.

Table 6 displays the models to predict the output for the three dependent variables previously defined in this paper under the leadership of Janet Yellen. In the first model, the prevailing independent variable is the S&P 500 growth rate, which is negatively impacting the outcome. The more the S&P 500 grew over the period and the more concise Yellen was likely to be. There was very little statistical significance in the other two models for positive and negative words. There was a slightly significant negative relationship between the Federal Funds rate and the use of positive words, which is consistent with expectations. There was also a slightly significant negative relationship with the unemployment rate and the usage of negative words. This is also consistent with expectations given Yellen's focus on employment during her tenure as chair of the FOMC.

CONCLUSION

The main objective of this paper is to examine the impact of various economic factors on the amount and type of language employed by the chair of the U.S. Federal Reserve. In doing so, it also assesses whether there are significant differences between the amount and type of language used by each Fed chairs and provides empirical evidence of how effectively this language interprets the economic condition of the US. The study shows there are statistically significant differences of language between each chairs of the U.S. Federal Reserve across time. Moreover, the communication strategy employed by each of the chairs is relying on a different combination of statistically significant independent economic variables. For instance, an interpretation of this empirical evidence is that the quantity of words employed in the Minutes by Greenspan is more greatly impacted by the unemployment rate and the federal funds rate, while for Yellen it is based upon the S&P 500 growth. The chairs of the U.S. Federal Reserve may therefore be representing the state of the economy in an empirically distinctive manner.

REFERENCES

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C. (2003). "Micro effects of macro announcements: real-time price discovery in foreign exchange", *American Economic Review*, Vol. 93, pp. 38–62.

Blinder, A., Ehrmann, M., Fratzscher, M., De Haan, J. and Jansen, D.-J. (2008). "Central Bank communication and monetary policy: a survey of theory and evidence", *Journal of Economic Literature*, Vol. 46, pp. 910–945.

Boukus, E. and Rosenberg, J. (2006). "The Information Content of FOMC Minutes", *Federal Reserve Bank of New York*, mimeo.

Campbell, J. R., Evans, C. L., Fisher, J. D. M. and Justiniano, A. (2012). "Macroeconomic Effects of FOMC Forward Guidance", *Brookings Papers on Economic Activity*, Spring 2012 conference.

Costa, B. A., & Jakob, K. (2011). "Are mutual fund managers selecting the right benchmark index?" *Financial Services Review*, 20(2), 129-143.

Sicilia, D. B., & Cruikshank, J. L. (2000). "The Greenspan effect: Words that move the world's markets." *New York: McGraw-Hill*.

Tausczik, Y. R., & Pennebaker, J. W. (2010). "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of Language and Social Psychology*, 22, 24-54.

ENDNOTES

[1] Greenspan's view of deflation was discussed in these articles:

https://www.wsj.com/articles/SB10358561966975251111

https://www.nytimes.com/2002/12/20/business/greenspan-s-speech-focuses-on-deflation-not-inflation.html

APPENDIX

Table 1: Descriptive Statistics

Panel A – Total Word Count

Chair	Sample Period	Sample Size (N)	Average	Standard Deviation	Minimum	Maximum
All	1993 - 2017	198	5871.21	2151.33	2851.00	13161.00
Greenspan	1993 - 2006	105	4702.50	1405.17	2851.00	9486.00
Bernanke	2006 - 2014	65	6624.06	1953.47	3336.00	11043.00
Yellen	2014 - 2017	28	8506.14	1800.50	6063.00	13161.00

Panel B – Positive Words

Chair	Sample Period	Sample Size (N)	Average (%)	Standard Deviation	Minimum	Maximum
All	1993 - 2017	198	2.30	0.33	1.30	3.22
Greenspan	1993 - 2006	105	2.18	0.31	1.30	3.00
Bernanke	2006 - 2014	65	2.49	0.32	1.69	3.22
Yellen	2014 - 2017	28	2.31	0.21	1.89	2.71

Panel C – Negative Words

Chair	Sample Period	Sample Size (N)	Average (%)	Standard Deviation	Minimum	Maximum
All	1993 - 2017	198	1.24	0.36	0.58	2.76
Greenspan	1993 - 2006	105	1.26	0.42	0.58	2.76
Bernanke	2006 - 2014	65	1.26	0.30	0.66	1.87
Yellen	2014 - 2017	28	1.13	0.20	0.81	1.59

Table 2: Correlation Matrix

Panel A: All Fed Chairs Correlation Matrix

	WC	*WC (%Max)*	*PosWord*	*NegWord*	*Real GDP Growth*	*S&P500 Growth*	*Inflation*	*Unemployment*	*FFR*	*3Mo T-Bill*	*6Mo T-Bill*	*12Mo T-Bill*	*3Yr T-Note*	*5Yr T-Note*	*10Yr T-Note*
WC	1***														
WC (%Max)	1***	1***													
PosWord	0.13590	0.13590	1***												
NegWord	-0.30080*	-0.30080*	0.19179	1***											
Real GDP Growth	-0.24558	-0.24558	-0.33381*	-0.22373	1***										
S&P500 Growth	-0.08095	-0.08095	-0.04098	-0.10446	0.22451	1***									
Inflation	-0.04597*	-0.04597	-0.07457	-0.16394	0.21049	0.04663	1***								
Unemployment	0.33408*	0.33408*	0.44739*	0.06176	-0.22415	0.02169	0.00061	1***							
FFR	-0.52856**	-0.52856**	-0.58783**	-0.18666	0.307333*	0.05545	0.12668	-0.63767**	1***						
3Mo T-Bill	-0.52856**	-0.52856**	-0.58783**	-0.18666	0.307333*	0.05545	0.12668	-0.63767**	1***	1***					
6Mo T-Bill	-0.52895**	-0.52895**	-0.59127**	-0.19216	0.306131*	0.05041	0.12571	-0.63688**	0.99851***	0.99851***	1***				
12Mo T-Bill	-0.53209**	-0.53209*	-0.59771**	-0.19397	0.314845*	0.04959	0.12591	-0.62775**	0.99371***	0.99371***	0.99757***	1***			
3Yr T-Note	-0.53892**	-0.53892**	-0.59792**	-0.18858	0.339045*	0.05035	0.13178	-0.58247**	0.96532***	0.96532***	0.97224***	0.98439***	1***		
5Yr T-Note	-0.54376**	-0.54376**	-0.58926**	-0.16716	0.33806*	0.04186	0.13773	-0.52691**	0.93336***	0.93336***	0.94113***	0.95803***	0.99215***	1***	
10Yr T-Note	-0.54579**	-0.54579**	-0.55633*	-0.13429	0.32463*	0.03565	0.15130	-0.40518*	0.87386***	0.87386***	0.88230***	0.90419***	0.95747***	0.98451***	1***

Panel B: Greenspan Correlation Matrix

	WC	*WC (%Max)*	*PosWord*	*NegWord*	*Real GDP Growth*	*S&P500 Growth*	*Inflation*	*Unemployment*	*FFR*	*3Mo T-Bill*	*6Mo T-Bill*	*12Mo T-Bill*	*3Yr T-Note*	*5Yr T-Note*	*10Yr T-Note*
WC	1***														
WC (%Max)	1***	1***													
PosWord	-0.35209*	-0.35209*	1***												
NegWord	-0.48730*	-0.48730*	0.33874*	1***											
Real GDP Growth	-0.10890	-0.10890	-0.12961	-0.20783	1***										
S&P500 Growth	0.00216	0.00216	-0.01541	-0.08983	0.14535	1***									
Inflation	0.11920	0.11920	0.08066	-0.24171	0.10498	-0.02089	1***								
Unemployment	0.14220	0.14220	0.03847	0.16972	-0.16274	-0.05573	-0.01539	1***							
FFR	0.205856307	0.205856307	-0.52533**	-0.47364*	0.151214238	0.093929437	0.025441993	-0.55432**	1***						
3Mo T-Bill	0.205856307	0.205856307	-0.52533**	-0.47364*	0.151214238	0.093929437	0.025441993	-0.55432**	1***	1***					
6Mo T-Bill	0.201799112	0.201799112	-0.53895**	-0.49694*	0.176567686	0.098127525	0.03866105	-0.53739**	0.99640***	0.99640***	1***				
12Mo T-Bill	0.21467674	0.21467674	-0.56214**	-0.51650**	0.193097116	0.098001572	0.041577182	-0.49034*	0.98402***	0.98402***	0.99419***	1***			
3Yr T-Note	0.25112294	0.25112294	-0.60364**	-0.52585**	0.197428257	0.089693841	0.044041149	-0.36599*	0.92777***	0.92777***	0.94799***	0.97415***	1***		
5Yr T-Note	0.284317257	0.284317257	-0.63431**	-0.49952*	0.173188634	0.076619319	0.038211946	-0.238151802	0.86337***	0.86337***	0.88764***	0.92566***	0.98533***	1***	
10Yr T-Note	0.31731*	0.31731*	-0.65231**	-0.45063*	0.146052335	0.059916528	0.036163616	-0.040948228	0.74169***	0.74169***	0.76992***	819193525***	0.91999***	0.97214***	1***

Panel C: Bernanke Correlation Matrix

	WC	*WC (%Max)*	*PosWord*	*NegWord*	*Real GDP Growth*	*S&P500 Growth*	*Inflation*	*Unemployment*	*FFR*	*3Mo T-Bill*	*6Mo T-Bill*	*12Mo T-Bill*	*3Yr T-Note*	*5Yr T-Note*	*10Yr T-Note*
WC	1***														
WC (%Max)	1***	1***													
PosWord	0.23107	0.23107	1***												
NegWord	-0.14244	-0.14244	0.01533	1***											
Real GDP Growth	-0.05564	-0.05564	-0.30295*	-0.35044*	1***										
S&P500 Growth	-0.09111	-0.09111	-0.07023	-0.18036	0.36968*	1***									
Inflation	0.11638	0.11638	-0.20025	-0.20652	0.31846*	0.07142	1***								
Unemployment	0.50827**	0.50827**	0.51573**	-0.01864	0.09159	0.14189	-0.00089	1***							
FFR	-0.64932**	-0.64932**	-0.57825**	-0.09369	0.07605	-0.02613	0.06436	-0.86114***	1***						
3Mo T-Bill	-0.64932**	-0.64932**	-0.57825**	-0.09369	0.07605	-0.02613	0.06436	-0.86114***	1***	1***					
6Mo T-Bill	-0.65685**	-0.65685**	-0.57082**	-0.06990	0.04987	-0.04725	0.04960	-0.86794***	0.99869***	0.99869***	1***				
12Mo T-Bill	-0.66126**	-0.66126**	-0.56018**	-0.06391	0.03834	-0.05356	0.04444	-0.86304***	0.99635***	0.99635***	0.99916***	1***			
3Yr T-Note	-0.64378**	-0.64378**	-0.51048**	-0.09933	0.03599	-0.04439	0.06575	-0.79859***	0.96936***	0.96936***	0.97499***	0.98175***	1***		
5Yr T-Note	-0.60312**	-0.60312**	-0.47383**	-0.11992	0.03632	-0.05770	0.08176	-0.73281***	0.92192***	0.92192***	0.92959***	0.93974***	0.98527***	1***	
10Yr T-Note	-0.53437**	-0.53437**	-0.41780**	-0.13825	0.04293	-0.06612	0.09680	-0.61562**	0.82595***	0.82595***	0.83562***	0.84925***	0.92498***	0.97542***	1***

Panel D: Yellen Correlation Matrix

	WC	*WC (%Max)*	*PosWord*	*NegWord*	*Real GDP Growth*	*S&P500 Growth*	*Inflation*	*Unemployment*	*FFR*	*3Mo T-Bill*	*6Mo T-Bill*	*12Mo T-Bill*	*3Yr T-Note*	*5Yr T-Note*	*10Yr T-Note*
WC	1***														
WC (%Max)	1***	1***													
PosWord	0.09369	0.09369	1***												
NegWord	-0.04901	-0.04901	-0.07444	1***											
Real GDP Growth	-0.16493	-0.16493	0.10343	-0.33736*	1***										
S&P500 Growth	-0.38199*	-0.38199*	0.17042	0.09036	-0.18975	1***									
Inflation	-0.17322	-0.17322	-0.15734	0.07486	0.07516	0.27258	1***								
Unemployment	-0.14769	-0.14769	0.07525	-0.29066	0.22086	-0.05544	-0.11041	1***							
FFR	-0.02460	-0.02460	-0.24973	0.02795	0.00235	-0.02794	0.13540	-0.78538***	1***						
3Mo T-Bill	-0.02460	-0.02460	-0.24973	0.02795	0.00235	-0.02794	0.13540	-0.78538***	1***	1***					
6Mo T-Bill	0.02498	0.02498	-0.24024	0.09537	-0.07999	-0.04500	0.12732	-0.84202***	0.98806***	0.98806***	1***				
12Mo T-Bill	0.04559	0.04559	-0.21284	0.15351	-0.14279	-0.01978	0.14225	-0.88820*	0.96601***	0.96601***	0.99087***	1***			
3Yr T-Note	-0.00992	-0.00992	-0.17600	0.00974	-0.12240	-0.09420	-0.01769	-0.72204***	0.85743***	0.85743***	0.86666***	0.86932***	1***		
5Yr T-Note	-0.26295	-0.26295	-0.11950	-0.23704	0.04771	-0.09718	-0.06571	-0.14197	0.48522***	0.48522***	0.44152*	0.40619*	0.76616	1***	
10Yr T-Note	-0.42019*	-0.42019*	-0.03846	-0.31272*	0.12231	-0.02842	-0.00527	0.39796*	0.01310	0.01310	-0.05628	-0.11224	0.27627	0.81825***	1***

Table 3: Regression Output for All Chairs (1993-2017)

Panel A – Dependent Variable = Percentage of Maximum Word Count

All Chairs	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	56.072***	29.797***	29.754***	30.017***	56.200***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.633	-0.140**	-1.154**	-1.237***	-0.634
	(0.170)	(0.022)	(0.018)	(0.009)	(0.150)
S&P 500 Growth	-0.134	-0.199	-0.199		
	(0.593)	(0.468)	(0.467)		
Inflation Rate	2.559	-0.577			
	(0.528)	(0.896)			
Unemployment Rate	-0.117	3.018***	3.013***	2.975***	-0.095
	(0.885)	(0.000)	(0.000)	(0.000)	(0.905)
FFR	-3.890***				-3.855***
	(0.000)				(0.000)
N	198	198	198	198	198
R^2	0.290	0.145	0.145	0.142	0.287

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel B – Dependent Variable = Positive Words

All Chairs	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	2.399***	1.908***	1.905***	1.905***	2.395***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.024***	-0.034***	-0.035***	-0.035***	-0.023***
	(0.006)	(0.000)	(0.000)	(0.000)	(0.006)
S&P 500 Growth	0.002	0.000	0.000		
	(0.722)	(0.926)	(0.924)		
Inflation Rate	0.027	-0.032			
	(0.728)	(0.704)			
Unemployment Rate	0.023	0.081***	0.081***	0.081***	0.024
	(0.145)	(0.000)	(0.000)	(0.000)	(0.121)
FFR	-0.073***				-0.072***
	(0.000)				(0.000)
N	198	198	198	198	198
R^2	0.380	0.258	0.258	0.258	0.379

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel C – Dependent Variable = Negative Words

All Chairs	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	1.517***	1.317***	1.304***	1.311***	1.534***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.023**	-0.027**	-0.031***	-0.033***	-0.028**
	(0.044)	(0.018)	(0.006)	(0.002)	(0.011)
S&P 500 Growth	-0.005	-0.005	-0.005		
	(0.454)	(0.411)	(0.418)		
Inflation Rate	-0.151	-0.175*			
	(0.136)	(0.084)			
Unemployment Rate	-0.019	0.005	0.004	0.003	-0.023
	(0.335)	(0.749)	(0.817)	(0.865)	(0.242)
FFR	-0.030*				-0.033**
	(0.058)				(0.034)
N	198	198	198	198	198
R^2	0.085	0.068	0.053	0.050	0.040

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 4: Regression Output for Greenspan (1993-2006)

Panel A – Dependent Variable = Percentage of Maximum Word Count

Alan Greenspan	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	1.793	27.352***	28.436***	28.488***	2.706
	(0.885)	(0.000)	(0.000)	(0.000)	(0.872)
Real GDP Growth	-0.717	-0.592	-0.511	-0.493	-0.641
	(0.179)	(0.296)	(0.366)	(0.375)	(0.222)
S&P 500 Growth	0.010	0.072	0.059		
	(0.967)	(0.782)	(0.821)		
Inflation Rate	6.295	6.542			
	(0.174)	(0.185)			
Unemployment Rate	4.692***	1.695	1.697	1.687	4.709***
	(0.002)	(0.198)	(0.199)	(0.199)	(0.002)
FFR	2.735***				2.749***
	(0.000)				(0.000)
N	105	105	105	105	105
R^2	0.166	0.045	0.028	0.028	0.150

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel B – Dependent Variable = Positive Words

Alan Greenspan	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	3.489***	2.189***	2.213***	2.213***	3.508***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.016	-0.23	-0.021	-0.021	-0.013
	(0.224)	(0.177)	(0.210)	(0.206)	(0.310)
S&P 500 Growth	0.004	0.000	0.000		
	(0.545)	(0.938)	(0.967)		
Inflation Rate	0.156	0.143			
	(0.182)	(0.327)			
Unemployment Rate	-0.146***	0.007	0.007	0.007	-0.145***
	(0.000)	(0.859)	(0.858)	(0.858)	(0.000)
FFR	-0.139***				-0.138***
	(0.000)				(0.000)
N	105	105	105	105	105
R^2	0.388	0.027	0.017	0.170	0.375

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel C – Dependent Variable = Negative Words

Alan Greenspan	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	2.354***	1.099***	1.027***	1.022***	2.295***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Real GDP Growth	-0.027	-0.033	-0.039*	-0.040*	-0.33*
	(0.153)	(0.118)	(0.075)	(0.060)	(0.085)
S&P 500 Growth	-0.004	-0.007	-0.006		
	(0.681)	(0.499)	(0.564)		
Inflation Rate	-0.422**	-0.434**			
	(0.011)	(0.020)			
Unemployment Rate	-0.076	0.071	0.071	0.072	-0.078
	(0.144)	(0.151)	(0.161)	(0.154)	(0.147)
FFR	-0.134***				-0.136***
	(0.000)				(0.000
N	105	105	105	105	105
R^2	0.307	0.115	0.065	0.062	0.259

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 5: Regression Output for Bernanke (2006-2014)

Panel A – Dependent Variable = Percentage of Maximum Word Count

Ben Bernanke	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	67.938***	19.709***	21.024***	21.962***	70.470***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	0.035	-0.564	-0.269	-0.559	0.149
	(0.945)	(0.396)	(0.671)	(0.348)	(0.788)
S&P 500 Growth	-0.326	-0.456	-0.482		
	(0.335)	(0.232)	(0.210)		
Inflation Rate	8.384*	7.680			
	(0.087)	(0.163)			
Unemployment Rate	-1.459	4.103***	4.073***	3.950***	-1.629
	(0.338)	(0.000)	(0.000)	(0.000)	(0.289)
FFR	-6.539***				-6.595
	(0.000)				(0.000)
N	65	65	65	65	65
R^2	0.470	0.311	0.288	0.269	0.432

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel B – Dependent Variable = Positive Words

Ben Bernanke	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	2.305***	1.885***	1.868***	1.870***	2.298***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.032**	-0.037***	-0.041***	-0.042***	-0.035***
	(0.024)	(0.008)	(0.002)	(0.001)	(0.005)
S&P 500 Growth	-0.001	-0.002	-0.001		
	(0.930)	(0.815)	(0.851)		
Inflation Rate	-0.097	-0.103			
	(0.379)	(0.356)			
Unemployment Rate	0.042	0.091***	0.091***	0.091***	0.042
	(0.220)	(0.000)	(0.000)	(0.000)	(0.219)
FFR	-0.057				-0.058*
	(0.111)				(0.098)
N	65	65	65	65	65
R^2	0.424	0.399	0.390	0.390	0.417

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel C – Dependent Variable = Negative Words

Ben Bernanke	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	1.565***	1.302***	1.284***	1.292***	1.573***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.029*	-0.033**	-0.037**	-0.039***	-0.03588
	(0.066)	(0.036)	(0.013)	(0.005)	(0.015)
S&P 500 Growth	-0.004	-0.004	-0.004		
	(0.669)	(0.609)	(0.638)		
Inflation Rate	-0.103	-0.107			
	(0.414)	(0.395)			
Unemployment Rate	-0.028	0.003	0.003	0.002	-0.030
	(0.486)	(0.885)	(0.868)	(0.910)	(0.436)
FFR	-0.036				-0.038
	(0.381)				(0.341)
N	65	65	65	65	65
R^2	0.148	0.137	0.126	0.123	0.136

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 6: Regression Output for Yellen (2014-2017)

Panel A – Dependent Variable = Percentage of Maximum Word Count

Janet Yellen	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	121.187***	85.099***	84.300***	80.235***	112.161***
	(0.003)	(0.000)	(0.000)	(0.001)	(0.007)
Real GDP Growth	-1.396	-2.005	-2.095	-1.328	-0.744
	(0.471)	(0.291)	(0.255)	(0.495)	(0.714)
S&P 500 Growth	-2.142**	-2.042**	-2.129**		
	(0.037)	(0.046)	(0.028)		
Inflation Rate	-2.546	-3.614			
	(0.831)	(0.763)			
Unemployment Rate	-9.102	-2.712	-2.553	-2.426	-8.102
	(0.179)	(0.499)	(0.512)	(0.566)	(0.261)
FFR	-14.901				-13.148
	(0.237)				(0.327)
N	28	28	28	28	28
R^2	0.271	0.222	0.219	0.040	0.079

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel B – Dependent Variable = Positive Words

Janet Yellen	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	3.011***	2.213***	2.164***	2.193***	3.058***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	0.038	0.024	0.19	0.013	0.029
	(0.226)	(0.433)	(0.542)	(0.658)	(0.335)
S&P 500 Growth	0.018	0.020	0.015		
	(0.251)	(0.214)	(0.338)		
Inflation Rate	-0.199	-0.223			
	(0.298)	(0.262)			
Unemployment Rate	-0.133	0.008	0.018	0.017	-0.136
	(0.216)	(0.897)	(0.779)	(0.789)	(0.201)
FFR	-0.329				-0.356*
	(0.106)				(0.078)
N	28	28	28	28	28
R^2	0.206	0.103	0.051	0.014	0.135

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Panel C – Dependent Variable = Negative Words

Janet Yellen	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	2.176***	1.558***	1.572***	1.576***	2.181***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Real GDP Growth	-0.031	-0.041	-0.040	-0.040	-0.029
	(0.293)	(0.157)	(0.159)	(0.138)	(0.290)
S&P 500 Growth	-0.002	0.000	0.002		
	(0.918)	(0.989)	(0.900)		
Inflation Rate	0.083	0.065			
	(0.644)	(0.723)			
Unemployment Rate	-0.176*	-0.066	-0.069	-0.069	-0.177*
	(0.089)	(0.280)	(0.248)	(0.237)	(0.074)
FFR	-0.255				-0.249
	(0.181)				(0.171)
N	28	28	28	28	28
R^2	0.235	0.168	0.163	0.163	0.227

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 7: Difference in Means – Total Word Counts

Panel A – Greenspan versus Bernanke

	Greenspan	**Bernanke**
Mean	4702.50	6624.06
Variance	1974501.60	3816037.12
Observations	105	65
Hypothesized Mean Difference	0	
df	105	
t Stat	-6.90	
P(T<=t) one-tail	<0.0001	
t Critical one-tail	1.66	
P(T<=t) two-tail	<0.0001	
t Critical two-tail	1.98	

Panel B – Greenspan versus Yellen

	Greenspan	**Yellen**
Mean	4702.50	8506.14
Variance	1974501.60	3241789.09
Observations	105	28
Hypothesized Mean Difference	0	
df	36	
t Stat	-10.37	
P(T<=t) one-tail	<0.0001	
t Critical one-tail	1.69	
P(T<=t) two-tail	<0.0001	
t Critical two-tail	2.03	

Panel C – Bernanke versus Yellen

	Bernanke	**Yellen**
Mean	6624.06	8506.14
Variance	3816037.12	3241789.09
Observations	65	28
Hypothesized Mean Difference	0	
df	55	
t Stat	-4.51	
P(T<=t) one-tail	<0.0001	
t Critical one-tail	1.67	
P(T<=t) two-tail	<0.0001	
t Critical two-tail	2.00	

Table 8: Difference in Means – Positive Words

Panel A – Greenspan versus Bernanke

	Greenspan	**Bernanke**
Mean	2.18	2.49
Variance	0.10	0.10
Observations	105	65
Hypothesized Mean Difference	0	
df	133	
t Stat	-6.17	
P(T<=t) one-tail	<0.0001	
t Critical one-tail	1.66	
P(T<=t) two-tail	<0.0001	
t Critical two-tail	1.98	

Panel B – Greenspan versus Yellen

	Greenspan	**Yellen**
Mean	2.18	2.31
Variance	0.10	0.04
Observations	105	28
Hypothesized Mean Difference	0	
df	63	
t Stat	-2.68	
P(T<=t) one-tail	0.000464	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.000927	
t Critical two-tail	2.00	

Panel C – Bernanke versus Yellen

	Bernanke	**Yellen**
Mean	2.49	2.31
Variance	0.10	0.04
Observations	65	28
Hypothesized Mean Difference	0	
df	77	
t Stat	3.15	
P(T<=t) one-tail	0.000116	
t Critical one-tail	1.66	
P(T<=t) two-tail	0.000231	
t Critical two-tail	1.99	

Table 9: Difference in Means – Negative Words

Panel A – Greenspan versus Bernanke

	Greenspan	**Bernanke**
Mean	1.264	1.262
Variance	0.17	0.09
Observations	105	65
Hypothesized Mean Difference	0	
df	163	
t Stat	0.03	
P(T<=t) one-tail	0.488527	
t Critical one-tail	1.65	
P(T<=t) two-tail	0.977054	
t Critical two-tail	1.97	

Panel B – Greenspan versus Yellen

	Greenspan	**Yellen**
Mean	1.26	1.13
Variance	0.17	0.04
Observations	105	28
Hypothesized Mean Difference	0	
df	92	
t Stat	2.50	
P(T<=t) one-tail	0.000709	
t Critical one-tail	1.66	
P(T<=t) two-tail	0.000142	
t Critical two-tail	1.99	

Panel C – Bernanke versus Yellen

	Bernanke	**Yellen**
Mean	1.26	1.13
Variance	0.09	0.04
Observations	65	28
Hypothesized Mean Difference	0	
df	76	
t Stat	2.56	
P(T<=t) one-tail	0.000619	
t Critical one-tail	1.67	
P(T<=t) two-tail	0.000124	
t Critical two-tail	1.99	

Crowding Out As a Cause of U.S. Declining Business Dynamism[1]

Chukwudi Ikwueze[2]

ABSTRACT

This paper discusses U.S. declining business dynamism. Specifically, it explores how U.S. exports and imports, the federal government deficit, and tax-exempt nonprofits contribute to U.S. declining business dynamism. Using data from Business Dynamics Statistics (BDS) of the U.S. Census Bureau, Master Files Databases of the Internal Revenue Service (IRS) and other secondary sources, the paper employs both co-integration and error correction models to perform the tests. The results suggest that government and the foreign sectors crowd out U.S. business formation.

1. INTRODUCTION

Theoretical analysis of the modern market economy has deep roots in the classical, neoclassical, and Austrian models (Simpeh, 2011) as well as the recent endogenous theorists such as (Brach, 2008). The classical economists hypothesized that it was the combination of private capital and property, the 'invisible hand' of the market and human labor that were the sources of economic growth (Smith, 1776). The neoclassical economists focus on physical and human capital accumulation as the ultimate sources of growth (Solow, 1956), and endogenous growth theorists champion technological progress and factor productivity (Solow, 1957; Arrow, 1962; Uzama, 1965).

A common characteristic of market economies has been the way firms are continually born, failing, expanding, and contracting (Schumpeter, 1942), a process referred to as business dynamism or creative destruction. The United States became the leading market economy due to the dynamic economic expansion recorded over the last two centuries. Over this period, U.S. business dynamism has been positive, meaning that more firms are born than fail, but there has been a noticeable decline, as described by Hathaway and Lithan (2014[a], p.1):

> The U.S. economy is in a constant state of churn. Historically one new business is born about every minute, while another one fails every eighty seconds. In 2012, there were 13.4 million private sector jobs created or destroyed each quarter—that's equivalent to one in eight private sector jobs. Despite all of that churning, only 600 thousand net jobs were created each quarter during that same year. That's equal to about half a percent of private employment.

Hathaway and Litan (2014[c]) show that U.S. declining business dynamism affects productivity and entrepreneurship, and results in consolidation of the monopoly power of older firms. Hathaway and Litan

[1] *Support for this project was provided by a PSC-CUNY Award, jointly funded by The Professional Staff Congress and The City University of New York.

[2] Department of Social Sciences, Queensborough Community College of the City University of New York. E-mail: cikwueze@qcc.cuny.edu

(2014[b]) and Hathaway et al. (2014) further show that U.S. business sector created 12 new firms per business establishment in 1978 to 6.2 in 2011. So, U.S. business dynamism has declined.

The question becomes what factors might be contributing to declining business dynamism in the United States? To address this question, this paper examines how U.S. exports and imports, the federal government deficit, and tax-exempt nonprofits may be contributing to declining U.S. business dynamism. Exports and imports are proxies for the effects of foreign competition on the U.S. economy. Intuitively, this may be used to measure the impact of globalization on the business sector. The federal government deficit is a proxy for the role of U.S. government in the economy. Tax-exempt nonprofits are included to find out if the increasing numbers of nonprofits in recent decades has encroached on the traditional activities of profit-making businesses. In other words, the federal government deficit and tax-exempt nonprofits may capture the impact of changing structure of the economy on the business sector.

This paper explores how these variables influence the level of business dynamism and is organized as follows. It explores the literature on business dynamism followed by a review to identify variables and sources of data. Then, the paper presents the model specification and test results and interpretations. In sum, we found that the government (federal government deficit) and foreign (compositions of export and import) sectors have impacted negatively on the U.S. business dynamism over the study period. The concluding remarks focus on the policy implications of the study findings.

2. A BRIEF LITERATURE REVIEW

Historically, the United States has experienced positive business dynamism, in which more businesses are created annually than fail. In recent years, however, available data suggest that the United States may be experiencing declining business dynamism, more businesses exit annually than are formed. The question then is why do businesses exit? The *selection or passive model* (Jovanovic, 1982) predicts that firms gradually learn about their relative abilities from the date of entry and then exit if they receive unfavorable information about the sustainability of their activities.

In the *active learning or evolutionary learning model* (Pakes and Ericson, 1992), the assertion is that firms exit if they lack the ability to improve (for example, by increasing productivity) and reduce the competitive gap between themselves and the incumbents. Therefore, both the selection/passive model and active learning/evolutionary learning model suggest that the exit rate declines with the age of the firm, which is positively correlated with the firm's productivity.

In contrast, the *vintage model* (Johansen, 1959, 1972; Solow, 1956, 1960) predicts that the exit rate of firms increases with the age of capital, where the age of capital is generally assumed to be positively correlated with the age of a firm. The main idea here is that new technology is embodied in the latest vintages of capital. Thus, new capital is better than old capital even when the old capital is new.

But, in arguing for both the selection and vintage models, Salvanes and Tveteras (2004) suggest incorporating both the firm's age and the age of the firm's capital where the latter is constructed by using investments in machinery based on Mairesse's (1978) approach to disentangling the distinct effects of

selection and vintage on exit rates. A U-shaped exit rate in the age of a firm is expected when both forces are active; the failure rate of a firm is expected to first decrease with the firm's age due to the selection and passive-learning model effects, and then to increase due to the vintage capital effect.

Still, Dunne (1994) criticizes the vintage model on the grounds that it has poor empirical support since old firms through investments often acquire the most recent technologies. Others (Davis and Haltiwanger, 1999; Caballero and Hammour, 1994, 1996) have also criticized the vintage model on the basis that a firm's age has been used as a proxy for the vintage of capital due to data limitations.

In the literature, there is also the *recession model* (Audretsch and Mahmood, 1995) that points out that the hazard rate of new firms increases with the unemployment rate, which is used as a proxy for the business cycle. In other words, new firms are more likely to fail during macroeconomic downturns.

But a study by Boeri and Bellmann (1995) finds that the hazard rates of new firms are not responsive to the business cycle and that growth of surviving entrants exhibits little cyclical sensitivity.

The *dynamic cost-of-adjustment model* asserts that an entering firm should experience high initial growth as it begins investing in capital but that the growth rate should fall as the firm completes its investment (Troske, 1996).

Other notable studies on why businesses fail include Christensen and Bower (1996), which finds that lack of adequate investment can lead to business failures; and Pfeffer and Salancik (1978) which finds that a firm fails because it may not invest in a new technology due to lack of demand for its products because purchasing power of existing customers influence the investment patterns of established firms. Tripsas (1997) shows that an established firm can be handicapped by its prior experience because its approach to new product development is shaped by that experience. For example, new products developed by established firms are likely to be inferior to those of new entrants (Cooper and Schendel, 1976; Majumdar, 1982; Tushman and Anderson, 1986; Afuah, 1994).

Furthermore, Gilbert and Newberry (1982) find that when innovation replaces rather than competes with old technology, existing firms have less incentive to invest in the new technology. But if innovation is incremental, existing firms have greater incentives than new entrants to invest; their findings suggest that existing firms exit when innovation replaces an old technology.

Another study by Thurik (2009) finds that even the shift from a *managed economy model* to an *entrepreneurial economy model* may contribute to business failures. In the managed-economy model, factors such as production, labor, and capital are dominant. This means that the more mobile capital moves to where the cheapest labor (software) is the more such labor moves towards capital once it is invested in plants (hardware).

In the entrepreneurial economy model, on the other hand, knowledge is the dominant factor of production. Here, knowledge includes not only hard technical and scientific knowledge but also soft aspects like creativity, the ability to communicate, and emotional intelligence.

Now, we review theories on why businesses are created. There appears to be a consensus In the literature that firms enter an industry if (excess) profits can be made under the conditions of no uncertainties

and opportunity cost to entry other than capital costs (Quandt, 1968; Conner, 1991). So, firms enter an industry when there are profits and exit when there are no profits. However, there are divergent views on what factors help to open up profits opportunity for firms. Kihlstrom and Laffont (1979) explore the entrepreneurial model of business formation to ultimately show that less risk-averse people become entrepreneurs while risk-averse people become workers. This means that risk-averse people are majorly the ones who create businesses and risk-averse people are those who become employees. The neoclassical (perfect competition) view asserts that firms are created if profits can be made to produce an end product by teaming labor and capital (Alchian, 1982; Alchian and Demsetz, 1972).

According to the resource-based model (Wernerfelt, 1984), firms are created because owners can mobilize tangible and intangible resources, that markets may not be able to access, to produce products that are distinctive in the eye of buyers or low-cost products relative to competitors.' Other notable resource-based studies conclude that knowledge (an intangible resource) determines the effectiveness of a business (Conner and Prahalad, 1996); knowledge integration rather than knowledge itself determines the effectiveness of a business (Grant, 1996); differences between firms in terms of resources and capabilities make a firm better than competitors (Peteraf, 1993); core competencies of a business, such as activities, knowledge, and internal organizational structure make a firm better than competitors (Hamel and Prahalad, 1990); and the availability of physical, financial, human, and organizatorial resources make a firm better than competitors (Barney, 1995).

Under the transaction cost model (Coase, 1937), firms are created to avoid incurring heavy transaction cost associated with going to the traditional market all the time. In this sense, a firm is a tool for reducing costs of operating in a market. This study finds also that firms may incur heavy transaction costs if they grow too big, explaining why large monopolies fail or become uncompetitive.

The opportunistic theory (Williamson, 1975, 1989) points out that firms are created due to opportunistic behavior which arises when three conditions occur simultaneously: asset specificity, small numbers of potential transactions and imperfect information. Opportunistic behavior is that which is dictated by unprincipled self-interest by taking advantage of asymmetry information for benefits (Lou et al., 2015). Klein and Leffler (1981), however, point out that opportunistic behavior is not sufficient to cause business formation; for example, when parties face the same magnitude of opportunistic behaviors, they tend to offset each other.

Bain-type view (Bain, 1954) holds that a firm enters an industry if it can acquire market power with which to control and deter competition. The point here is that firms may not get created if they cannot acquire some market power with which to control and deter competition.

Schumpeter's view (Schumpeter, 1950) is that a firm enters an industry if it can seize competitive opportunity through creation and adoption of innovation that make rivals obsolete. The Chicago view (Stigler, 1968) posits that a firm enters an industry if it can produce and distribute products more efficiently than rivals through the market and price mechanisms.

But the above theories of why firms exit and enter an industry may not fully explain why U.S. business

dynamism has declined. This is because the processes of exit and entry of firms, as captured by the theories, have been properly researched, and there is no evidence that U.S firms have abandoned free-enterprise practices implied by the theories, such as private-property system, cost minimization, profit motives, and maximizing shareholders equity. So, what then has caused the processes of exit and entry of firms in the United States suddenly to begin to generate unsatisfactory outcomes?

Thus, we have to look beyond traditional theories in search for reasons that business dynamism has worsened in the United States. In this study, we attempt to identify major periodic events that have affected U.S. business sector over the study period, basing our reasoning on the theoretical premise that an economy is prone to the effects of periodic events (Rogers, 1982; Ikwueze, 2014). We focus our investigation on two events that occurred or grew in significance over this period: globalization (Parry, 2004) and the changing structure of U.S. economy (Kehoe et al., 2013). Therefore, we investigate the impact on U.S. declining business dynamism of the composition of exports and imports (as proxies for rising foreign competition due to globalization), the federal government deficit, and increasing number of tax-exempt nonprofit organizations (as proxies for changing structure of the U.S. economy).

3. VARIABLES AND SOURCES OF DATA

The dependent variable (BUSDYNUS) in this study is the annual gap between U.S. business formation/entry and failure/exit rates. There are four explanatory variables: exports (EXPORTUS), imports (IMPORTUS), the federal government deficit (DEFICITUS), and numbers of tax-exempt nonprofits (TAXXMTUS).

BUSDYNUS: The dependent variable represents the annual percentage rate of business entry-exit gap in the United States from 1981 to 2014. The data for BUSDYNUS are collected from the Business Dynamics Statistics (BDS) database of the U.S. Census Bureau.

EXPORTUS: This explanatory variable represents the annual dollar value of U.S. exports from 1981 to 2014. The data for EXPORTUS are collected from the World Trade Organization (WTO) database for time series merchandise trade with the world.

IMPORTUS: This explanatory variable represents the annual dollar value of U.S. imports from 1981 to 2014. The data for IMPORTUS are collected from the World Trade Organization (WTO) database for time series merchandise trade with the world.

DEFICITUS: This explanatory variable represents the annual U.S. federal government deficit as percentage of GDP from 1981 to 2014. The data for DEFICITUS are collected from a secondary source, USgovernment.us/spending.

TAXXMPTUS: This explanatory variable represents the numbers of organizations that received a tax exemption from the Internal Revenue Services (IRS) from 1981 to 2014. Overall, the reasoning behind selecting this variable is that some nonprofits are allowed to run, for benefit of their members, bakeries, skeletal financial services, construction, food handling and so on. So, as the number of nonprofits has grown, the concern is that their business-related activities may have encroached on the traditional activities of the

business sector. The data for TAXXMPTUS are collected from IRS Master File databases. The variables and sources of data are shown in Table 1 and graphed in Figures 1 and 2.

TABLE 1: Variables and Data Sources

YEAR	BUSDYNUS[a]	EXPORTUS[b]	IMPORTUS[c]	DEFICITUS[d]	TAXXMTUS[e]
1981	-8	238715	273352	2.46	851012
1982	3.9	216442	254884	3.83	841440
1983	1.1	205639	269878	5.71	845464
1984	3.5	223976	346363	4.59	871224
1985	3.1	218815	352463	4.88	886658
1986	3.1	227158	382294	4.82	929415
1987	3.5	254122	424443	3.07	978676
1988	1.7	322427	459543	2.95	1012365
1989	1.0	363812	492922	2.70	1038070
1990	3.0	393592	516987	3.70	1073443
1991	2.0	421730	508363	4.36	1107664
1992	1.5	448163	553923	4.44	1140388
1993	1.2	464773	603438	3.71	1177772
1994	2 .0	512627	689215	2.78	1203238
1995	2.3	584743	770852	2.14	1235905
1996	1.7	625073	822025	1.33	1266802
1997	1.8	689182	899020	0.25	1322505
1998	1.4	682138	944353	-0.76	1376395
1999	1.2	695797	1059440	-1.30	1428208
2000	.7	781918	1259300	-2.30	1473062
2001	.6	729100	1179180	-1.21	1567580
2002	.9	696103	1200230	1.44	1580767
2003	2.7	724771	1303050	3.28	1640949
2004	2.3	814875	1525680	3.36	1680061
2005	2.9	901082	1732706	2.43	1709205
2006	2.3	1025967	1918077	1.79	1726491
2007	1.5	1148199	2020403	1.11	1789554
2008	.1	1287442	2169487	3.12	1855067
2009	-2.9	1056043	1605296	9.80	1912695
2010	-.7	1278495	1969184	8.65	1960203
2011	.4	1482508	2266024	8.37	1629149
2012	1	1545703	2336524	6.73	1616053
2013	.7	1579593	2329060	4.08	1599013
2014	1.4	1620532	2412547	2.79	1723315

Source: Prepared by the author.

[a] U.S. Census Bureau [Business Dynamics Statistics]: Annual differences between U.S. business entry and exit

[b] World Trade Organization [Statistics Database]: Volume of Exports in million dollars

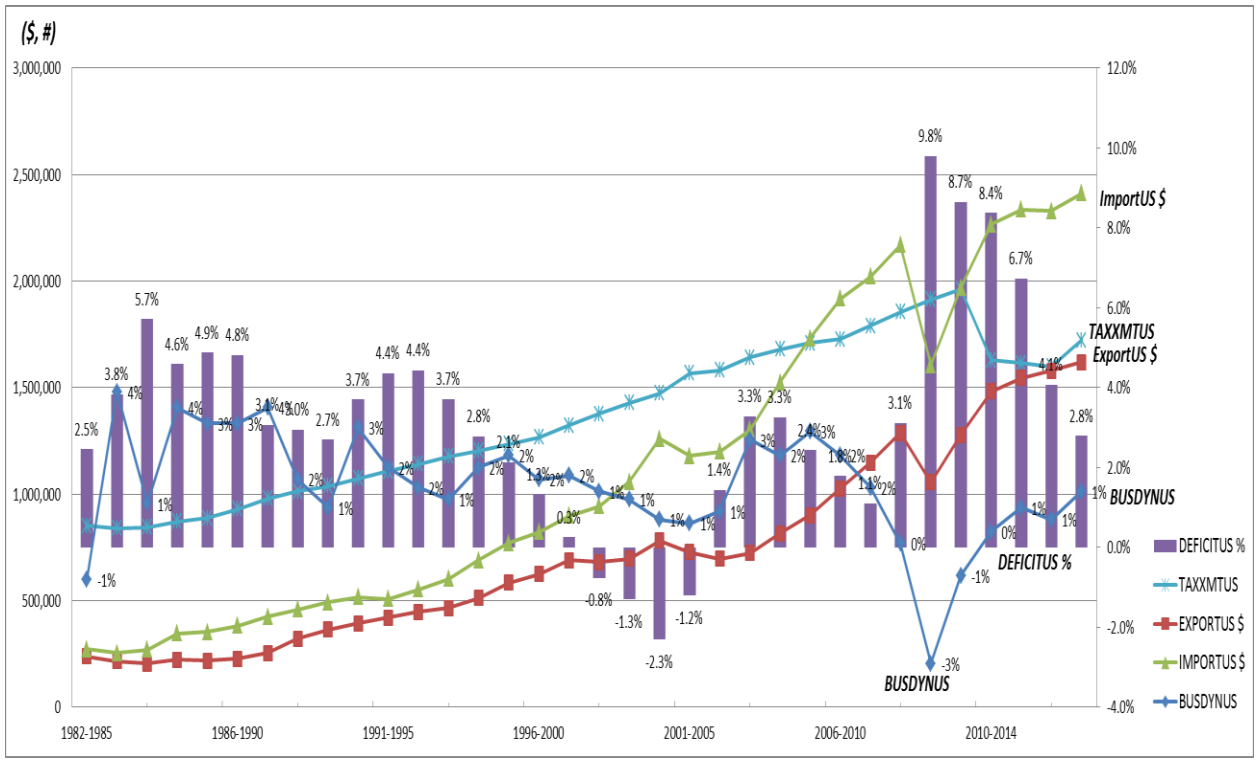
[c] World Trade Organization [Statistics Database]: Volume of Imports in million dollars

[d] USgovernment.us/spending [online]: Annual U.S. Federal Government Deficits as percent of GDP

[e] Internal Revenue Service [Master File Databases]: Annual Numbers of Tax-Exempt Nonprofits

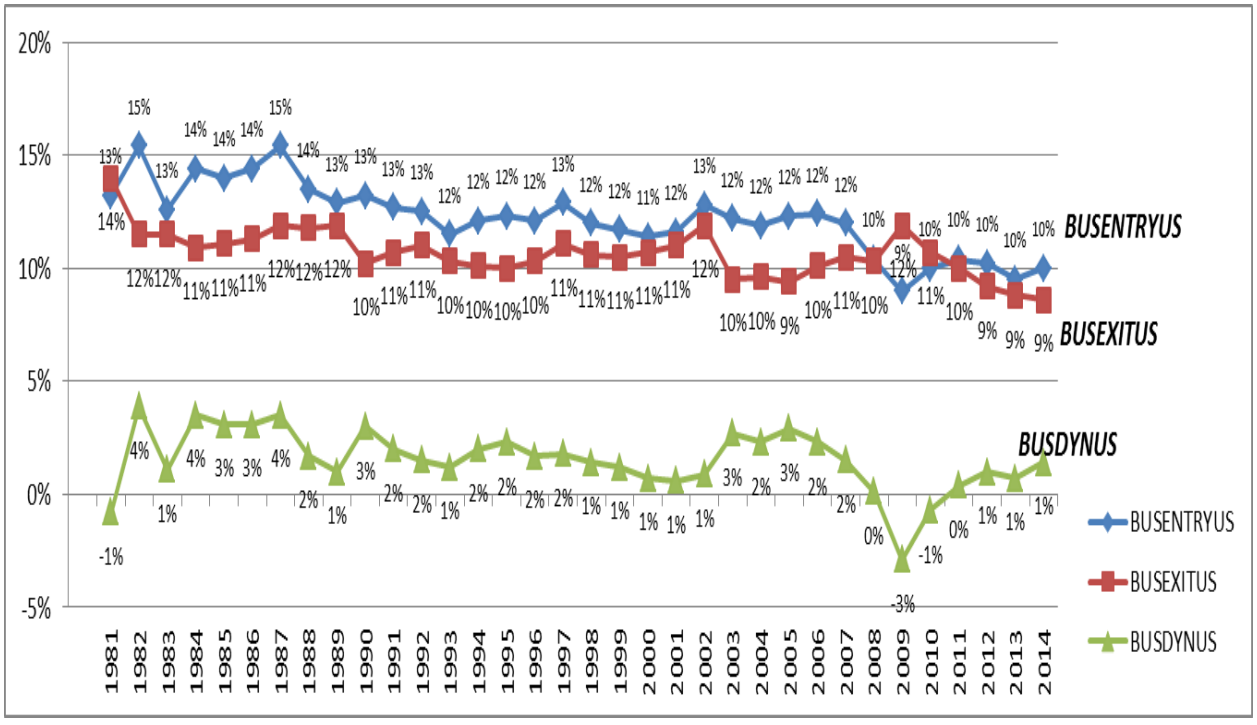
Figure 1 shows the four explanatory variables (DeficitUS, ExportUS, ImportUS, TaxxmtUS) and the dependent (BusdynUS) variable from 1981 to 2014, based on the data in Table 1. DeficitUS was high over the period, but there was a surplus from 1997 to 2001. TaxxmtUS showed that the number of tax-exempt nonprofits increased. ExportUS and ImportUS showed that per year imports exceed exports and trade deficit has increased over time. BusdynUS, which captures the percentage annual gap between business entry and exit, has declined during the period.

Figure 1: U.S. Business Dynamism, Deficit, Tax-Exempt Nonprofits, Exports, and Imports (1981-2014)



Source: Prepared by the author.

Figure 2: U.S. Business Entry, Business Exit, and Business Dynamism (1981-2014)



Source: Prepared by the author.

Figure 2 shows annual percentage change in business entry (BusEntryUS), Exit (BusExitUS), and BusDynUS (annual percentage difference between business entry and exit) from 1981 to 2014. Notice that BusdynUS did not rise above three percent and two percent after 1987 and 2006, respectively. Clearly, this confirmed that U.S. business dynamism narrowed/declined over the period. We now introduce the theoretical framework of the models for analyzing the data.

4. MODEL SPECIFICATION

An autoregressive distributed lag model (ADL (1,1)) in its simplest form is used (Chen, 2010):

$$Y_t = m + A_1 Y_{t-1} + B_o X_t + B_1 X_{t-1} + U_t \quad \textbf{(1)}$$

where Y_t (dependent variables in year t) and X_t (independent variables in year t) are stationary variables, and U_t is white noise. A sequence $\{U_t\}$ is a white noise process if each value in the sequence has a mean of zero, constant variance, and is serially uncorrelated. Equation 1 can be rewritten as follows:

$$\Delta Y_t = m + B_0 \Delta X_t - (1-A_1) Y_{t-1} + (B_0 + B_1) X_{t-1} + U_t \quad \textbf{(2)}$$

We further rearrange Equation 2:

$$\Delta Y_t = B_0 \Delta X_t - (1-A_1)[Y_{t-1} - (m/(1-A_1)) - ((B_0+B_1)/(1-A_1))X_{t-1}] + U_t . \quad \textbf{(3)}$$

Equation 3 is an error correction model (ECM). One of the vital features of the ECM is that it enables researchers to establish short-run and long-run relationships between the dependent and independent variables. So, if the ECM results show that the variables have long-run relationships, because policies usually take a long time to work, then it is an indication that any policies, based on the results, would have enough time to work. For the purposes of this study, therefore, if there were long-run relationships between the gap in U.S. business entry and exit rates on the one hand and the explanatory variables on the other, then, we could interpret the tests' results to mean that policies formulated to reverse the declining U.S. business dynamism would have enough time to work. Consequently, that would also mean that explanatory variables that fail to show a long-run relationship to the dependent variable, gap in business entry and exit rates, may be hazardous to policymaking (Shittu et al., 2012). We now perform a series of tests to determine the relationships between the dependent and explanatory variables.

5. TESTS AND INTERPRETATION OF RESULTS

5.1 Unit Root Test

The Dickey Fuller (DF) unit root test is one of the most commonly used tests for stationarity (Dickey and Fuller, 1981). The standard Augmented Dicker-Fuller test is performed to assess the degree of integration of selected variables. The null hypothesis is that the time-series data have a unit root: $H:^{w} 0$, and the alternative hypothesis is that the time-series data are stationary: $H:^{w} 1$. One way to deal with it is to include lagged values of the dependent variable in the DF regression. An autoregressive model can be derived (Said and Dickey, 1984) in the general drift/trend case:

$$\Delta Y_t = B_0 + B_0 Y_t + \delta t + (A-1) Y_{t-1} + \Sigma B_t \Delta Y_{t-1} + \varepsilon_t . \quad \textbf{(4)}$$

where ε_t is independently and identically distributed, i.i.d $(0,\sigma^2)$.

Table 2 presents the results of the Augmented Dicker-Fuller tests.

TABLE 2: Results of ADF Stationarity Test of Variables

	Level		First Difference (FD)		
	ADF		ADF		
Variables	**(Intercept)**	**(Trend and Intercept)**	**(Intercept)**	**(Trend and Intercept)**	**Decision**
BUSDYNUS*	-2.277 (-2.980)	-2.744 (-3.572)	-4.040 (-2.983)	-3.967 (-3.576)	Accept I(1)
DEFICITUS	-1.843 (-2.980)	-2.426 (-3.572)	-3.669 (-2.983)	-3.633 (-3.576)	Accept I(1)
EXPORTUS	1.125 (-2.980)	-1.580 (-3.572)	-4.737 (-2.983)	-5.416 (-3.576)	Accept I(1)
IMPORTUS	0.423 (-2.980)	-0.914 (-3-576)	-5.218 (-2.983)	-5.429 (-3.576)	Accept I(1)
TAXXMTUS	-1.139 (-2.980)	-1.894 (-3.572)	-3.413 (-2.983)	-3.478 (-3.576)	FDIntercept: Accept I(1) FDIntercept+Trend: I(0)

Source: Computed by the author. *Figures in parenthesis are critical value at 5% and the corresponding figures are the t-statistic values. We used Lag 1 to conduct the tests.

Table 2 shows that the dependent variable of BUSDYNUS and three explanatory variables of DEFICITUS, EXPORTUS, and IMPORTUS are not stationary, meaning potentially that the variance and autocorrelation structure change, over time, at the first difference and are integrated of order one. So, while the variables are non-stationary, their first differences appear to be stationary, suggesting the levels are integrated of order 1, I(1). We shall now perform co-integration tests to determine the features of the stationary long-term relationships between the variables.

5.2 CO-INTEGRATION TEST

Co-integration analysis helps to test for the existence of a stationary long-run relationship between variables. The basic idea is that although two or more variables are non-stable series, but some linear combination of them would be stable (Engel and Granger, 1987). In that sense, there is co-integration among variables, implying long-term and stable relationship. Table 3 presents the results of the co-integration tests:

TABLE 3: Co-Integration Johansen Test Results

Maximum Rank	Parms	LL	Eigenvalue	Trace Statistic	5% Critical Value
0	30	-1255.3033	.	123.8617	68.52
1	39	-1225.9644	0.84018	65.1841	47.21
2	46	-1209.2025	0.64923	31.6601	29.68
3	51	-1197.7719	0.51052	8.7989*	15.41
4	54	-1193.4776	0.23539	0.2104	3.76
5	55	-1193.3724	0.00655		

Source: Computed by the author. We used Lag 2 to conduct the tests.

In Table 3 there are at least three co-integrating equations where the trace statistic is larger than the critical value. Therefore, the explanatory variables of DEFICITUS, EXPORTUS, IMPORTUS, and possibly TAXXMPTUS co-integrate with the dependent variable of BUSDYNUS. This means that the variables, together, have long-run relationships. We interpret the tests' results to mean that even though the variables are not stationary, as shown in Table 2, there are at least three long-run stationary relationships linking them together (Escudero, 2000). As referenced elsewhere, above, and for purposes of reiteration, it is worth noting that the main reason we want to find out if there are long-term relationships between the variables is to make sure that any policies formulated to reverse U.S. declining business dynamism would have enough time to work. We now estimate an error correction model to determine the magnitude and direction of the stationary linkages between the variables. In Table 4, the error correction estimation results are shown:

TABLE 4: Vector Error Correction Model

AIC = 78.04831 **HQIC = 78.82264** **SBIC = 80.38433**

Det (Sigma_ml) = 2.23e+26 **Log Likelihood = -1197.773**

D_BUSDYNUS	COEF	STD. ERR.	Z	P>|Z|	[95% Confidence Interval]	
ECT|	-.7230361	.2980285	-2.43	0.015	-1.307161	-.138911
EXPORTUS|	-.0000205	5.61e-06	-3.65	0.000	-.0000315	-9.50e-06
IMPORTUS|	. 8.30e-06	3.52e-06	2.36	0.018	1.40e-06	.0000152
DEFICITUS|	-.4185282	.1913681	-2.19	0.029	-.7936028	-.0434536
TAXXMTUS|	-5.57e-06	3.09e-06	-1.80	0.072	-.0000116	4.95e-07
CONS|	1.529138	.8045818	1.90	0.057	-.0478137	3.106089

Source: Computed by the author. Lagrange-multiplier tests: 96 percent; Jarque-Bera tests: p-value of seventy-nine percent for all variables combined. We used Lag 2 to conduct the tests.

Table 4 shows the results of using the error correction model test to determine the relationships between U.S declining business dynamism (BUSDYNUS) and the compositions of U.S. exports, imports, federal government deficits, and tax-exempt nonprofits. The result error correction term (ECT) is -.7230361, which means that less than seventy-two percent of the previous year's disparity in the U.S. business dynamism can potentially be restored in the current year. Also, because the estimated ECT model's p-value is 0.015, (which is significant), and both the coefficient and z statistic have negative signs, there are long-run relationships among the variables. Furthermore, notice in Table 4, notice that EXPORTUS, IMPORTUS, and DEFICITUS have p-values of 0.000, 0.018, and 0.029, respectively. This means that the three explanatory variables have significant short-run relationships with U.S. declining business dynamism (BUSDYNUS). However, Table 4 also shows that TAXXMTUS has a p-value of 0.057, which shows that TAXXMTUS has no significant short-run explanatory relationships with U.S. declining business dynamism (BUSDYNUS).

To ensure that the data for this study are not biased for autocorrelation, we further perform Lagrange-multiplier tests. This test enables researchers to measure the level of test, which is the probability of rejecting a suitable model. For example, when five percent is used, it indicates that there is a five percent chance of rejecting a study's model by mistake. Thus, the Lagrange-multiplier test can be used to reject an unsuitable model. For this study's model, as shown in the footnote of Table 4, the p-value for Lag 2 is ninety-six percent. Given that this p-value is above five percent, it suggests that the null hypothesis on each test is acceptable, implying that there is no autocorrelation in the series. Then, we check for the residual's normality using the Jarque-Bera tests. The results show a p-value of more than five percent for each variable, and for all variables, when considered jointly, it is approximately seventy-nine percent (Lag 2), as shown in the footnote of Table 4. This means that the residuals are normally distributed, so all tests' results are acceptable.

6. CONCLUDING REMARKS

This study focused on finding out how the current composition of U.S. exports, imports, government deficits, and the number of tax-exempt nonprofits affected U.S. business dynamism from 1981 to 2014. The results of the co-integration and error correction tests showed that the four variables, jointly have long-run associations with U.S. declining business dynamism. The value of the model's error correction term of ---0.230361, means that approximately seventy-two percent of the disparity or gap in business dynamism from the previous year may be corrected within a year. In addition, apart from tax-exempt nonprofits' variable, the results showed that the federal government deficit, exports, and imports have short-run associations with the U.S. declining business dynamism. Thus, our findings clearly suggest that the government (federal government deficit) and foreign (compositions of export and import) sectors had a negative impact on U.S. business dynamism over the last three and half decades. By implication, this means that the government and foreign sectors are crowding out the U.S. business sector. According to this study's findings, this crowding out phenomenon has been caused by globalization and changing structure of the U.S. economy that began or accelerated in significance over the study period.

We suggest therefore that the United States consider directing policies to reduce the growing government deficit and encourage more exports than imports (leading to a positive net export balance) in order to achieve net business formation. This may be achieved in in the following ways, for example, given the state of physical infrastructure like railroads, airport, and roads, government could maintain or slightly increase the level of expenditure on infrastructural improvements if it could reduce corporate tax burdens enough to trigger the repatriation of more than two trillion five hundred billion dollars belonging to U.S. corporations in foreign banks (Cox, 2016). If these policies were implemented, the United States could potentially experience a boom in business activities that may end up affecting positively the levels of government deficit and net export, but also help forestall the U.S. declining business dynamism.

ACKNOWLEDGEMENT

The author is grateful to the unanimous reviewer for valuable comments and to W. O'Dea for his editorial assistance and comments. Also, I would like to thank participants at the annual conference of the New York Economics Association at Sienna College, Albany, New York, 2015.

REFERENCES

Afuah, A. N. 1994. Strategic adoption of innovation: The case of reduced instruction set computer technology. (Ph.D. Dissertation, Massachusetts Institute of Technology, Massachusetts, United States of America).

Alchian, A.A.1982. Property rights, specialization, and the firm. In J. F. Weston and M.E. Granfield (Eds.). Corporate Enterprise in a New Environment: 11-36. New York: KCG Production.

Alchian, A.A. and Demsetz, H. 1972. "Production, Information Costs, and Economic Organization," *American Economic Review*, 62: 777-794.

Arrow, K. 1962. "The economic implications of learning by doing," *The Review of Economic Studies,* 29(3): 155-173

Audretsch, D., and Mahmood, T. 1995. "New firm survival: New results using a hazard function," *Review of Economics and Statistics*, 97: 97-103.

Bain, J. S. 1954. "Economies of Scale, Concentration, and Condition of Entry in Twenty Manufacturing Industries, *American Economic Review*, 44: 15-39.

Barney, J. 1995. "Looking Inside for Competitive Advantage," *Academy of Management Executive*, 9(4): 49-61.

Boeri, T., and Bellmann, L. 1995. "Post-entry behavior and the cycle: Evidence from Germany," *International Journal of Industrial Organization*, 13: 483-500.

Brach, J. 2008, September. "Constraints to economic development and growth in the Middle East and North Africa," *German Institute of Global and Area Studies* (Working Paper No.85, p.9). April: giga-hamburg.de/en/system/files/publications/wp85_brach.pdf

Caballero, R., and Hammour, M. 1994. “The cleansing effect of recessions,” *American Economic Review*, 84: 1350-1368.

Caballero, R., and Hammour, M. 1996. “On the timing and efficiency of creative destruction,” *Quarterly Journal of Economics*, 111: 805-852.

Chen Yi Yi. 2010. Autoregressive distributed lag model. April. http://mail.tku.edu.tw/chenyiyi/ADL.pdf

Christensen, C., and Bower, J. 1996. “Customer power, strategic investment, and the failure of leading firms,” *Strategic Management Journal*, 17(3): 197-218.

Coase, R. 1937. “The Nature of the Firm,” Economica, New Series, Vol.4, No.16: 386-405.

Conner, K. 1991. “A Historical Comparison of Resource-Based Theory, and Five Schools of Thought Within Industrial Organization Economics: Do We Have a New Theory of the Firm,” *Journal of Mana*gement, Vol. 17, No. 1: 121-154.

Conner, K., and Prahalad, C. K. 1996. “A Resource-Based Theory Theory of the Firm: Knowledge Versus Opportunism,” *Organizationa*l *Science*, Vol. 7, Issue 51: 477-501.

Cooper, A.C., and Schendel, D. 1976. “Strategic responses to technological threats,” *Business Horizon*: 61-69.

Cox, J. 2016, September 20. “U.S. companies are hoarding $2.5 trillion dollars in cash overseas.” *CNBC Finance*. April. http://www.cnbc.com/2016/09/20/us-companies-are-hoarding-2-and-a-half-trillion-dollars-in-cash-overseas.html

Davis, S.J., and Haltiwanger, J. 1999. “On the driving forces behind cyclical movements in employment and job reallocation,” *American Economic Review*, Vol.89, No.5: 1234-1258.

Dickey, D. A., and Fuller, W.A. 1981. “Likelihood ratio statistics for autoregressive time series with a unit root,” *Econometrics*, Vol.49, No.4.

Dunne, T. 1994. “Plant age and technology use in U.S. manufacturing industries,” *Rand Journal of Economics*, Vol.25, No.3: 488-499.

Engle, R.F., and Granger, W.J. 1987. “Co-integration and error correction: Representation, estimation, and testing,” *Econometrica*, Vol. 55, No. 2: 251-276.

Escudero, W.S. 2000. “A Primer on Unit Root and Cointegration.” *Econometrea,* pp.1-17.

Gilbert, R., and Newberry, D. 1982. “Preemptive patenting and the persistence of monopoly,” *American Economic Review*, 72(2): 514-526.

Grant, R.M. 1996. “Prospering in a Dynamically Competitive Environment: Organizational Capabilities as Knowledge Integration,” *Organizational Science*, 7(4): 375-387.

Hamel, G., and Prahalad, C. K. 1990. “The Core Competence of the Corporation,” *Harvard Business Review*, pp. 79-90.

Hathaway, I., and Litan, R.E. 2014[a]. “Declining business dynamism in the United States: A look at states and metros,” *Economic Studie*s, Brookings Institution, p.1. http://www.brookings.edu/~/media/research/files/papers/2014/05/declining_business_dynamism_hathaway_litan.pdf

Hathaway, I., and Litan, R.E. 2014[b]. "The other aging of America: The increasing dominance of older firms," *Economic Studies*, Brookings Institution, p.9. http://www.brookings.edu/research/papers/2014/07/aging-america-increasing-dominance-older- firms-litan

Hathaway, I., and Litan, R.E. 2014[c]. "What's driving the decline in firm formation rate? A partial explanation," *Economic Studies*, Brookings Institution, p.9. http://www.brookings.edu/research/papers/2014/11/driving-decline-firm-formation-rate-hathaway-litan

Hathaway, I., Schweitzer, M.E., and Shane, S. 2014. "The shifting source of new business establishments and new jobs," *Economic Commentary*, Federal Reserve Bank of Cleveland, pp.14. http://www.clevelandfed.org/research/commentary/2014/2014-15.pdf

Ikwueze, L.C. 2014. "Cohort Analysis of U.S. Consumer Bankruptcy," *New York Economic Review,* Vol. XLV, pp. 47-66.

Johansen, L. 1959. "Substitution versus fixed production coefficients in the theory of economic Growth," *Econometrica*, 27: 157-176.

Johansen, L. 1972. *Production functions: An integration of micro and macro, short-run and long-run aspects.* North-Holland: Amsterdam, London.

Jovanovic, B. 1982. "Selection and the evolution of industry," *Econometrica*, Vol. 50, Issue 3: 649-670.

Kehoe, T.J., Ruhl, K.J., Steinberg, J.B. 2013. "Global Imbalances an the Structural Change in the United States," Federal Reserve Bank of Minneapolis. Research Department Staff Report 489, pp.1-44.

Kihlstrom, R.E., Laffont, J. 1979. "A General Equilibrium Entrepreneurial Theory of Risk Aversion," *Journal of Politica*l Ec*onomy,* Vol. 87, No.4: 719-748.

Klein, B. and Leffler, K. B.1981. "The Role of Market Forces in Assuming Contractual Performance," *Journal of Politica*l Ec*onomy,* 89: 615-641.

Lou, J., Meyers, J., and Dignum. 2015. "On Formalizing Opportunism-Based on Situation Calculus." In F. Koch et al. (Eds.): CARE-MFSC 2015, CCIS 541, pp.147-166. Doi: 10.1007/978-3-319-24804-2-10.

Mairesse, J. 1978. "New estimates of embodied and disembodied technical progress," *Annales de l'Insee*, 30/31: 681-720.

Majumdar, B.A. 1982. *Innovations, product developments and technology transfers: An empirical study of dynamic competitive advantage. The case of electronic calculators*, Washington, D.C.: University Press of America.

Pakes, A., and Ericson, R. 1992. "Empirical implications of an alternative model of firm dynamics," NBER Working Paper No. 2893.

Parry, R.T. 2004. "Globalization: Threat or Opportunity for the United States?" *FRBSF Economic Letter*. Federal Reserve Bank of San Francisco. Issue 21.

Peteraf, M.A.1993. "The Cornerstones of Competitive Advantage: A Resource-Based View," Strategic Management *Journal, 14: 179-191.*

Pfeffer, J., and Salancik, G.R. 1978. *The external control of organizations: A resource dependence perspective.* New York: Harper and Row.

Quandt, R., and Howrey, E.P. 1968. "The Dynamics of the Number of Firms in an Industry," *Review of Economic Studies*, 35: 349-353.

Rodgers, W. L. 1982. "Estimable Functions of Age, Period, and Cohort Effects," *American Sociological Review,* 47(6): 774–787.

Said, S.E., and Dickey, D.A. 1984. "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71: 599-607.

Salvanes, K. G., and Tveteras, R. 2004. "Plant exit, vintage capital and the business cycle," *The Journal of Industrial Economics*, 52: 255–276. doi: 10.1111/j.0022-1821.2004.00225.x

Schumpeter, J. A. 1942. *Capitalism, socialism, and democracy*. New York: Harper Bros**.**

Schumpeter, J. A. 1950. *Capitalism, socialism, and democracy* (3rd Ed.). New York: Harper and Row.

Shittu, O.I., Yemitan, R.A., and Yaya, O.S. 2012. "On autoregressive distributed lag, co-integration and error correction model: An application to some Nigeria macroeconomic variables," *Australian Journal of Business and Management Research*, Vol.2, No.08, p.57.

Simpeh, K.N. 2011. "Entrepreneurship theories and empirical research: A Summary Review of the Literature" *European Journal of Business and Management*, Vol.3, No.6, pp.1-2.

Smith, A. 1776. An inquiry into the nature and causes of the wealth of nations. Retrieved from http://www.adamsmith.org/smith/won-intro.htm

Solow, R.M. 1956. "A Contribution to the Theory of Economic Growth. *Quaterly Journal of Economics*, 70, No.1: 65-94.

Solow, R.M. 1957 "Technical change and aggregate production function," *The Review of Economics and Statistics*, Vol. 39, No.3: 312-320.

Solow, R.M. 1960. Investment and technical progress. In Arrow, K, Karlin, S., & Suppes, P. (Eds.) *Mathematical methods in the social science* (pp. 89-104) Stanford, CA: Stanford University Press.

Stigler, G. J. 1968. The economies of Information. In G. J. Stigler, *The organization of industry*: 171-190. Homewood: Irwin.

Thurik, R. 2009. "Entrepreneurship, economic growth and policy in emerging economies," *World Institute for Development Economics Research* (Research paper No. 2009/12).

Tripsas, M. 1997. "Unraveling the process of creative destruction: Complimentary assets and incumbent survival in the typesetter industry," *Strategic Management Journal*, Vol.18: 119-142.

Troske, K.R. 1996. "The dynamic adjustment process of firm entry and exit in manufacturing and finance, insurance, and real estate," *Journal of Law and Economics*, Vol.39, No.2: 705-735.

Tushman, M.L. and Anderson, P. 1986. "Technological discontinuities and organizational Environments," *Administrative Science Quarterly*, 31: 431-465.

U.S. Census Bureau. Data Tables: Establishment characteristics-economywide, business dynamics statistics. April. http://www.census.gov/ces/dataproducts/bds/data_estab.html

Uzama, H. 1965. “Optimum Technical Change in an Aggregate Model of Economic Growth,” *International Economic Review* 6(1): 18-31.

Wernerfelt, B.1984. “A Resource-Based View of the Firm,” Vol. 5, Issue 2: 171-180.

Williamson, O.E.1989. Transaction Cost Economics. In R. Schmalensee and R. D. Willig (Eds.), *Handbook of Industrial Organization*: 136-182. Amsterdam: North Holland.

Williamson, O.E.1975. Markets and Hierarchies: Analysis and Antitrust Implications. New York: Free Press.

World Trade Organization. Statistics: merchandise trade and commercial services Database. April. www.wsDBviewData.aspx?Language=E

REFEREE ACKNOWLEDGMENT

The New York State Economics Association is grateful for the time spent by the following referees in recent years. Without your efforts, we would not be able to produce this journal.

1. Kasia Platt
2. Julie Anna Golebiewski
3. Renzo Alvarez
4. Richard Deitz
5. Sean MacDonald
6. Alexander McQuoid
7. Michael Toma
8. Brent Evans
9. Shahdad Nagshpour
10. Eddery Lam
11. Garrison Song
12. Pinar Uysal
13. Paul Bauer
14. James Booker
15. Fan Chen
16. Kai Chen
17. Robert Jones
18. James Murtagh
19. Philip Sirianni
20. Jeffrey Wagner
21. Mark Zaparowski
22. Xu Zhang