Business Cycles: A Role for Limit Pricing and Countercyclical Markups in the Banking System

Federico Mandelman[1]

Boston College

MARCH 30, 2005

Abstract

This paper studies the cyclical pattern of banking system ex-post markups using annual aggregate bank data for a large set of countries for the period 1990-2001. Dynamic panel estimates show that markups are strongly countercyclical, even after controlling for financial development, banking concentration, operational costs, inflation, and simultaneity or reversed causation.

The countercyclical pattern is explained by entry of foreign banks, that occurs mostly at wholesale level, and signals the intention to spread later to some of the retail niches. However, the pro-competitive effect of entry in the local banking system is short lasting and vanishes after one year. One possible explanation is that booming periods lead to an expansion of the financial system that attracts potential competitors working at an efficient scale even at the retail level. In this situation, contestable markets force incumbent banks to charge markups well below short-run profit maximizing levels so as to protect their retail niches in a highly segmented banking system.

In the second part of the paper I develop a general equilibrium model that accounts for these features of the data. I find that this monopolistic behavior in the intermediary financial sector increases the volatility of real variables and amplifies the business cycle. I interpret this "bank lending channel" as an extension of the "credit channel" pioneered by Bernanke and Blinder (1988).

---

[1]E-mail: mandelma@bc.edu. I thank my advisors Fabio Ghironi, Peter Ireland and Fabio Schiantarelli for their guidance. Stephen Bond, Hideo Konishi, Todd Prono and seminar participants at R@BC and Boston College Dissertation Workshops provided very useful suggestions. The usual disclaimer applies.

1 Introduction

Under the standard assumption of factorial diminishing returns, Real Business Cycle models imply that at high levels of output, marginal products and factorial retributions are low. However, this implication does not square with the facts, real wages are procyclical in reality. The most familiar explanation for this puzzling prediction follows from the recognition of monopolistic goods markets that behave more competitively during booms. As a result, output prices fall relative to marginal costs and markups fall relative to wages and factor contributions. Essentially, markups are countercyclical. Ample empirical and theoretical evidence in goods markets provides support to this idea.[2] Instead, this paper will be concerned on financial markets where practically no work in the matter exists. Essentially, the questions this paper will address are:

1- If a countercyclical pattern in the financial markups also exists, what generates it?

2-What are the implications of this pattern in the real economy?

Limit Pricing strategies will be the answer this paper will propose for the first question. Limit pricing is the practice of setting prices at the limit level that deters entry. As it is shown in Bain (1956), the price level in an industry strongly influences firms contemplating entry. Thus, temporary low prices are not the result of changes in the market structure but just reflect the optimal strategy for the incumbent. In this scenario, the threat of entry is the only reason to avoid profit maximization.

Foreign penetration in the banking system has a wide effect in this matter. In the last decade, banks have expanded internationally by establishing foreign subsidiaries and branches or by taking over established banks. The internalization of the banking sector has been spurred by the liberalization of financial markets worldwide. Moreover, it is observed foreign bank penetration commonly takes place in the wholesale banking market initially and then expands to some of the retail level (Claessens et al 2001). Therefore, we can predict that the threat of foreign banks encroaching on retail markets

[2] For instance, see Pigou(1927), Phelps and Winter(1970), Rotemberg and Saloner(1986), Greenwald et al (1984), Murphy et al (1989), Bils (1989), Galí (1994), Chevallier and Sharfstein (1996), Galeotti and Schiantarelli (1998). Finally, nominal rigidities also generate countercyclical markups in standard new keynesian setups with Calvo pricing assumptions.

may also induce greater efficiency of established banks at the retail level.

However, the penetration into the retail sector is obstructed for the requirement of incurring in large sunk entry costs. For instance, large advertisement expenditures aimed to gain reputation or the construction of a network of branches and ATMs required to carrying small size transactions. This implies that banks need to enter at minimum-efficient-scale (MES) in order to justify the sunk costs incurred. In turns, it follows that right after entering they must capture a significant fraction of the market to make the constructed net workable. Something particularly difficult in the bank industry where within the country the markets are segmented in regional or sectorial niches (Rajan and Petersen, 1994). In this scenario, the size of the market constitutes a barrier to entry. If the financial market is small or underdeveloped there is "space" only for a few number of incumbents working at an efficient scale. Thus,

- Booming periods leads to an expansion of the financial system that attracts potential competitors with the possibility of working at an efficient scale. In this situation, contestable markets force incumbents to charge markups well below short run profit maximizing levels so as to avoid entry.

- Contrary, during recessions, the actors in the local financial system are able to exert their monopolistic power charging high markups.

This behavior not only would explain this pattern of financial markups but also will provide support to some well documented facts, like: a) Bank spreads being more countercyclical in concentrated markets. b) Immediate efficiency gains right after deregulation and much before any change in the market structure c) Ambiguous and contradicting relationship between concentration and efficiency. That is, my hypothesis predicts that incumbents combine periods of monopolistic markups with periods in which the efficiency gains from consolidation (and exploitation of economies of scale and scope) prevail.

Toward answering the second question, I rely on evidence portraying an important role for financial development in the magnitude of the business cycle [3]. If effectively these markups are countercyclical, we may be in presence of a "bank lending channel" that could be an extension of the "credit channel". In essence, the two channels would contribute independently to the same vicious circle: Credit is more expensive during recessions, and firms

[3] For a Survey, see Gertler and Hubbard (1988).

and households postpone investment and work decisions pushing the recession further. But while the latter channel relies on the external finance problem that induces the bank to charge a premium to cover the increasing expected bankruptcy costs during recessions, the former will be the solely result of imperfect competition in the banking system. This channel maybe particular relevant in developing countries, where bank credit remains the primary source of funds for entrepreneurs.

To test the validity of this hypothesis I will use bank data across 124 countries for the years 1991-2000, a decade with abundant liberalization and market deregulation processes across the world.

I will use Dynamic Panel Techniques to confront the potential bias induced by simultaneity and examine whether the exogenous component of the business cycle can explain the fluctuation in the ex-post markups. Since past work shows that long run-economic growth is a good predictor of financial development, I will attempt to remove the effect of the long-run trend component of the GDP growth series by controlling for a three year average of financial development.

To assess the strength of an independent link between the markups and the business cycle, I will use various conditioning sets that will include a proxy for concentration, overhead costs (operative and administrative costs), inflation volatility and changes in real interest rates.

Nonetheless, the cyclical behavior of the markups vanishes when controlling for entry of foreign banks. I interpret that entry at a wholesale level signals the intention to spread later to some of the retail niches. This evident threat triggers limit pricing aimed to influence entrant beliefs of the profitability of entering into a particular niche. In order to discard the possibility of foreign entry effect on markups being the result of a transformed (i.e. less concentrated and efficient) market structure (rather than the result of strategic incumbents' reaction), I will make a sensitivity analysis to show that the competitive pressure of entry is short lasting.

As an extension, I will check if the reaction to entry is different in the banking systems of developed and developing countries. I would expect a greater reaction in the in the latter group of countries, since the reach of financial development is restricted, and subject to a more bothersome regulation. The last section of this paper will propose a Dynamic Stochastic General Equilibrium model aimed to provide a grasp of the macro implications of the limit pricing scheme and the resulting "bank lending channel." The microfoundation of the novel limit pricing setup will account for several of the features of the data.

The paper will be organized as follows. In the following section I will proceed with a literature review. In section 3, I will present the data. In section 4, I will discuss the methodology to be used. In section 5, I will present the empirical results. Section 6 will introduce the theoretical model. Concluding remarks will be presented in section 7.

2 Literature review

The first step is to find a proper measure for markups in the banking industry data. In the existent literature, a simple approach is to consider the ex-ante spreads, or the difference between lending and deposits rate, as a proxy for financial markups. The difficulty here is that the spreads also include a premium to cover the expected borrowers' bankruptcy costs that increase during recessions and causes the spread to be countercyclical. To overcome this computation problems, I will use net interest margins (NIMs') as a proxy to measure markups.[4] They are measured as bank's total interest income minus interest expense over total assets. As explained in Demirguc-Kunt and Huizinga (1998), bank interest margins can be seen as an indicator of the inefficiency of the banking system as they drive a wedge between the interest rate received by savers on their deposits and the effective interest rate paid by borrowers on their loans. I will not consider the actual posted interest rates, but only ex-post measures. In the Data set, the price of banks assets and the deposit rates are taken directly from the balance sheets items. In other words, the proposed measure is based on actual income obtained by the banks after accounting for defaulted loans. Thus, net interest margins capture the pricing behavior and allows us not to be necessarily concerned with controlling for risk in our estimation analysis.[5]

[4] Angelini and Cetorelli (2000), developed an innovative strategy aimed to obtain estimates of Lerner indexes that measure the relative markup of price over marginal costs. There are however some drawbacks for this approach. The first one is purely technical: to obtain the Lerner indexes is necessary to count with disaggregated bank level data. Besides, the estimation of the cost function for the individuals banks requires "prices" for all the production factors (e.g. deposits, labor and capital). These data are rarely available for most of the countries studied here.

[5] One weakness of the method proposed is the treatment of income from services, which has become increasingly relevant in recent years. However, this is a not necessarily a

Practically all the existent literature is focused on ex-ante spreads.[6] An exception is Angellini and Cetorelli (2003) that considers the growth of GDP as an additional control variable in the estimation of interest margins and Lerner indexes for the Italian banking industry, finding a negative association. However, they do not settle the issue of causality and endogeneity. They also show that net interest margins remained relatively constant until 1992, declining after a wide process of financial liberalization and consolidation albeit with an increase in 1995 in coincidence with a monetary policy tightening. Similarly, in the U.S., Hannan and Berger (1991) find that after a monetary contraction ex-ante spreads tend to increase more in regional markets where the banking industry is more concentrated.

Although markups cyclicality did not deserve authors particular attention, there is an enormous literature on bank structure and efficiency. Actually, the literature is populated by ambiguous results. In a survey Rhoades (1977) expresses "disbelief and frustration" in the overall inability to link concentration and efficiency. New surveys and studies reach the same conclusions.[7] Contradictory results must be preceded by contradictory theories. The intuition of the Structure-Conduct-Performance (S-C-P) hypothesis is straightforward: a more concentrated market lowers the costs of collusion

concern here. I will not be interested in the profitability of the banks *per se* but in the disintermediation generated between borrowers and depositors. Strictly speaking margins are not markups over *total* marginal costs. That is, in them we just include the *cost of funds* for the banks but not the *operational costs*. However, I will control for these overhead costs in the regressions and make this distinction in the theoretical model.

[6] For instance, Olivero (2004) measures ex-ante spreads for the U.S. economy, and finds that GDP growth per capita is a good predictor of the spread values. The relationship is negative, even after controlling for the number of delinquent loans. Using a VAR analysis, Edwards and Vegh (1997) conclude that a world interest rate shock has its greatest negative impact on the level economic activity but also results in an increment of the ex-ante spreads in Chile and Mexico.

[7] For a new survey, see Bank for International Settlements 2001. Some empirical evidence shows that banks in highly concentrated local market have larger overhead costs expenditures, charge higher rates on loans, pay lower rates on deposits and are slower to reduce rates in response to Federal Reserve reductions in interest rates than banks in relatively less concentrated markets (See, for instance, Berger and Hannan (1989) and Neumark and Sharpe (1992)). However, others disagree. Smirlock (1985) and Grady and Kyle (1979) find that interest rate spreads are, instead, largely narrower in concentrated banking systems. Finally, Keeley and Zimmerman (1985) report mixed results. At a worldwide level the results do not differ. In a cross section analysis over 1,400 banks across 72 countries, Demirguc-Kunt et al (2003) found, at the same time, high net interest margins associated with both small banks and banks with a large market share.

and foster tacit or explicit collusion on the part of the banks. On the other hand, the efficient-structure (E-S) hypothesis predicts efficiency gains from market consolidation. Firms possessing a comparative advantage in production become large and, as a natural consequence, the market becomes more concentrated. Such cost differences maybe due to differences in technological or managerial skills. The effect is amplified because of large economies of scale existent in the bank industry. They are derived from risk diversification, lower average administrative costs (Demsetz 1973); and the efficient use of already incurred large sunk costs like the construction of large networks of branches and ATM's (Cerasi et al, 1997). Additionally, Gilligan et al (1984) have provided evidence that banking is characterized by economies of scope from joint production of financial services. Finally, in the absence of restrictions on entry, excessive inefficient profits are precluded (Baumol, 1982). These conclusions leaded to a new sort of literature aimed to find evidence of efficiency gains resulting from mergers and acquisitions. The results are, once again, ambiguous and inconclusive.[8]

Different are the results regarding bank deregulation and efficiency which are mostly unambiguous and conclusive. Several surveys hold that new legislations that remove substantial entry barriers and expose national banking markets to potential new entrants produces pro-competitive effect and reduce margins. Besides, banks loose market power following financial liberalization even in cases in which the banking industry remains highly concentrated.[9]

A more interesting result about deregulation is related with the timing of the efficiency gains. In a systematic study over 80 countries, Claessens et al (2001) find that foreign bank entry is significantly associated with a reduction in domestic bank profitability. However, the impact of foreign bank entry on local bank competition is "..felt immediately upon entry decision is taken rather after they have gained substantial market share". Angelini and Cetorelli (2000) find net interest margins declining sharply immediately after the banking reform is made effective in Italy. Similarly, Shaffer (1993) analyzes the impact of Bank Act Revisions in Canada and finds an already

[8]See for instance, Focarelli et al 1999, Prager and Hannan, 1999; Simons and Stavins, 1998 and Petersen and Rajan 1994.

[9]For a survey see Vives (1991) and Demirguc-Kunt (2003). Also Spiller and Favaro (1987) focus on the pro-competitive impact of the relaxation of entry restrictions in the Uruguayan industry, concluding that collusive strategic interactions across banks significantly decrease after the regulatory reform. Ribon and Yosha (1999) reach similar conclusions for the Israeli banking industry.

perfectly competitive conduct prior to the liberalization and evidence of an "unexpected supercompetitive state" right after, with negative interest margins observed. The author concluded that such atypical outcome ".. is not consistent with long run equilibrium behavior under known static or dynamic models of profit maximization; and it may simply reflect a temporary disequilibrium...(which) may warrant further study". These results resemble my hypothesis of limit pricing.

3 Data

I will construct an unbalanced panel from several data sources. The resulting comprehensive sample covers 124 countries during the years 1991-2000. It includes all the OECD countries, as well many developing countries and economies in transition. Bank structure information is taken from the database of indicators of financial development and structure published by the World Bank Research Department.[10] This information is originally compiled from the Scope Database provided by IBCA which contains data for 137 countries. To ensure reasonable coverage, only countries with at least three banks in a given year are included. Coverage by IBCA is comprehensive, roughly accounting for 90% of the assets of banks in each country. Each country has its own data template which allows for differences in account conventions. However, these are converted to a global format which is a globally standardized template derived from the country-specific templates. In the regressions, I will control for unobserved time invariant country specific effects so as to tackle minor differences regarding the valuation of assets that will necessarily remain.

Measures of activity of financial intermediaries are taken from the Levine-Loayza-Beck Data Set. Macroeconomic Data comes from the Penn World Table 6.1 (PWT 6.1.). Data on real interest rates is taken from the World Development Indicators 2002. Institutional data is taken from both the International Country Risk Guide and Dollar and Kraay (2001). Variable Definitions are provided in the appendix. I also show some few descriptive statistics. The degree of financial development captured by the ratio *Private Credit/ GDP* is significantly larger in developed countries than in developing ones. Consequently, net interest margins are significantly higher in the latter

[10] I will make use of both the 1999 and 2003 editions.

group. That is, 571 bp for these ones and 268 bp for the developed ones. Finally, poorer countries have a relatively high degree of concentration and foreign penetration.[11]

4 Econometric Issues

Panel data techniques are required in a context where the focus of the study consists on the cyclical pattern of the variables. While the aggregation of time series would obscure underlying microeconomic dynamics, Panel Data allows the investigation of heterogeneity in adjustment dynamics between different types of individuals.

The estimation procedure needs to tackle some important issues. Firstly, I must permit for the presence of unobserved country specific effects that are correlated with the regressors. Secondly, most of the explanatory variables in the specifications to be used (e.g. GDP growth rates, private credit, etc.) are most likely to be determined jointly with the dependent variable (i.e. net interest margins); therefore, I must also allow and control for joint endogeneity. Third, I will have to allow for the presence of non-time-varying variables. Fourth, I will need to use a dynamic specification in order to allow for inertia in the dependent variable, very likely to be present in the annual balance sheet information that will be used. Hereafter, I describe the techniques proposed to solve these problems.

The first dynamic specification will consist on a simple AR(1) model:

$$y_{it} = \alpha y_{i,t-1} + (\eta_i + \varepsilon_{it}) \qquad \mid \alpha \mid < 1 \qquad i = 1, 2, ..., N; \qquad t = 2, 3, ..., T. \quad (1)$$

[11]Notice that bank-level disaggregated data is not available in this study. I must rely on aggregate data for each country over a sample period of few years. There are other topics that will not be adressed in the paper because of data constraints. In Olivero (2004) the reason behind observed counter-cyclical markups is, actually, cyclical semi-elasticity of demand. Nonetheless, Focarelli and Rossi (1998) estimate demand schedules for bank credit in Italy and do not find evidence of elasticity instability in the recent period. Finally, it is expected that in period of exceptional banking crises revenues will decline, reducing bank margins. However, banking status must be linked to the economic cycle. Since I am proposing that margins actually increase during recessions, the omission of such negative bias would not put into question the validity of the hypothesis but the contrary.

The variable in consideration is net interest margin (NIM), and y_{it} denotes an observation for country i in period t; η_i is an unobserved individual-specific time-invariant effect and ε_{it} is a disturbance term that is assumed to be independent across individuals.[12] Since I treat the individual effects, η_i, as being stochastic; they are necessarily correlated with the lagged dependent variable, $y_{i,t-1}$. I further assume that the disturbances are serially uncorrelated. These jointly imply that the OLS estimator of α is inconsistent.

The Within Groups estimator eliminates this source of inconsistency by obtaining mean values across the $T-1$ observations in order to remove η_i. Nonetheless, for panels (like this one) where the number of time periods available is small, this transformation induces a non-negligible negative correlation between the transformed lagged dependent variable $y_{i,t-1}-\frac{1}{T-1}(y_{i1}+..+y_{it}+...+y_{i,T-1})$ and the transformed error term $\varepsilon_{i,t-1}-\frac{1}{T-1}(\varepsilon_{i2}+..+\varepsilon_{i,t-1}+...+y_{i,T})$. Standard results for omitted variable bias indicate that the OLS estimator is biased upwards and the Within Groups one is biased downwards. Therefore, a consistent candidate estimator must lie between the OLS and Within Groups estimates. I will exploit this fact.

The first-difference transformation of (1) also eliminates η_i from the model, but the dependence of $\Delta\varepsilon_{it}=\varepsilon_{it}-\varepsilon_{i,t-1}$ on $\varepsilon_{i,t-1}$ implies that OLS estimates are inconsistent. Nonetheless, consistent estimates of α can be obtained using two step least square estimations (2SLS) with instruments that are both correlated with $\Delta y_{i,t-1}$ and orthogonal to $\Delta\varepsilon_{it}$. The GMM estimator is asymptotically efficient since the set of all available instruments is used.[13] Specifically, the GMM *difference* specification is obtained from the following moment conditions:

$$E\left[y_{i,t-s}\Delta\varepsilon_{it}\right]=0 \qquad \text{for } s\geq 2;\ t=3,...,T. \tag{2}$$

Under the homoskedasticity assumption of the disturbances, $\Delta\varepsilon_i$,we can construct a *one-step* estimator based on a weighting matrix that does not

[12] For robustness, time dummies will also be included in order to account for time-specific effects.

[13] Maximum Likelihood Estimators for the AR(1) exists, but with the awkward characteristic that different assumptions about the nature of the initial conditions, y_{i1}, will lead to different likelihood functions, that will result in inconsistent estimators for α if the initial condition is mispecified. The shorter are the time length of the panel, the more serious in the problem.

depend on any estimated parameters. Otherwise we can proceed in *two-steps* and use consistent estimates of the first differenced residuals previously obtained from a preliminary consistent estimator. Since there is no a clear preference between these two estimators in the applied work literature, I will provide results for both. [14]

If $T > 3$ the model is overidentified, and the validity of the assumptions used in the estimation can be tested using the standard GMM Sargan test of overidentifying restrictions (under the null that these moment conditions are valid). The key identifying assumption that there is no serial correlation in the ε_{it} disturbances can also be tested. If the pattern of serial correlation in the first-differenced disturbances is consistent; $\Delta\varepsilon_{it}$ should have significative negative first-order serial correlation but not significant second-order serial correlation.

The *difference* GMM estimators for autoregressive models outlined before can be extended to models that also include a vector of current and lagged values of additional explanatory variables x_{it}. Since x_{it} is assumed to be endogenous, it is treated symmetrically with the dependent variable y_{it}. In this case, the lagged values $x_{i,t-2}, x_{i,t-3}$ and longer lags will be valid instruments in the first-differenced equations for periods $t = 3, 4, ..., T$. The proposed panel estimator controls for endogeneity by using "internal" instruments (i.e. instruments based on lagged values of the explanatory variables.) Specifically, all the explanatory variables are assumed to be weakly exogenous. This means that the explanatory variables are uncorrelated with future realizations of the error term and thus are not affected by future realizations of the dependent variable. The explanatory variables, however may be affected by current and past realizations of the dependent variable. This assumption permits for the possibility of simultaneity and reverse causality.

There are, however, several and serious econometric shortcomings with the *difference estimator* in the presence of inertia in the dependent variable. In particular, if the lagged dependent variable is persistent over time (i.e. near unit root), lagged levels of the variables are weak instruments for the

[14]Simulation studies have suggested very little gains from using the *two-steps* version even in the presence of heteroskedasticity. Besides, the dependence of the weighting matrix on estimated parameters makes the usual asymptotic distribution approximation less reliable particularly for small samples. See Blundell, Bond and Windmeijer(2000). On the other hand, Windmeijer (2000) proposes a solution for this problem by using a finite-sample correction for the asymptotic variance of the two step GMM estimator I will use.

regressions in differences.[15] To improve upon and solve this concerns, Blundell and Bond (1997) propose an alternative *system estimator*, that combines the regression in differences with the regression in levels. The instruments and moment conditions for the regression in differences are the same as above.

For the second part of the system (the regression in levels) the instruments are given by the lagged differences of the corresponding variables. These instruments are valid when the stationarity assumption is suitable.[16] In this scenario, the resultant non-redundant linear moment conditions for the second part of the system are:

$$\begin{aligned} E\left[\Delta y_{i,t-1}(\eta_i+\varepsilon_{it})\right] &= 0 \qquad i=1,2,...,N; \qquad t=3,4,...,T. \\ E\left[\Delta x_{i,t}(\eta_i+\varepsilon_{it})\right] &= 0 \qquad i=1,2,...,N; \qquad t=2,3,...,T. \end{aligned} \tag{3}$$

It exists the possibility of including non-time varying predetermined covariates in the mentioned set of explanatory variables. Institutional indicators can be assumed to possess such characteristic in the short span of time considered in this study. There are $T-1$ non-redundant moment conditions that allow for the computation of estimates despite the implausibility of considering equation for differences. These can be written as:

$$E\left[x_i(\eta_i+\varepsilon_{it})\right]=0 \qquad i=1,2,...,N; \qquad t=2,3,...,T. \tag{4}$$

[15]The issue is that parameters may not be identified using first-differenced GMM estimators when the series are random walks. Other serious problems exist. First, differences of the explanatory variables are often less correlated over time than levels. As noted by Barro (1997), this may produce biased estimates if the dynamic structure of the differenced equation model differs from the true model. Simulation studies show that the difference estimator has a large finite-sample bias and poor precision, particularly in samples with a small time series-dimension. Finally, by first differencing we end up loosing cross-country dimension and exploit only the time series dimension within countries.

[16]Under this assumption there might be a correlation between η_i and the levels of the variables, but this correlation is constant over time. That is, $E\left[y_{i,t+p}.\eta_i\right]=E\left[y_{i,t+q}.\eta_i\right]$ *for all p and q* and $E\left[x_{i,t+p}.\eta_i\right]=E\left[x_{i,t+q}.\eta_i\right]$ *for all p and q*. The validity of the sationarity assumptions about the initial conditions y_{i1} requires $E\left\{\left[y_{i1}-\left(\frac{\eta_i}{1-\alpha}\right)\right]\eta_i\right\}=0$ $for\ i=1,...,N$

The use of the described “full instrument set” in either the system and difference estimator result in the number of moment conditions tested growing rapidly as T increases. It is shown, that the size properties of the Sargan test are less sensitive to the number of moment conditions becoming large for a given cross-sectional sample size N (Bowsher, 2000). Although, it does not necessarily imply that the GMM estimator is biased or the standard errors are unreliable, I have to test whether the use of excessive number of instruments can made the estimator itself subject to a serious overfitting bias. The usual way to proceed is either restricting the set of explanatory variables, or if necessary to consider only the three closest lags to the regression period (for each variable) as instruments.[17] Finally, if the results of an equivalent system estimator based in a combination of orthogonal deviations and level equations significantly differs from the standard system; it may indicate that small biases are important. For robustness I will report both.[18]

5 Empirical Results

Time Series Properties The first two columns of Table 1 report OLS levels and Within Group estimates of the parameter α, jointly with heteroskedasticity-consistent estimates of the asymptotic standard errors. Recall that OLS estimate is likely to be biased upwards and the Within

[17] As it is repeatedly remarked, the loss of relevant information caused by omitting the more distant lags as instruments will often be very modest (Bond, 2002). For robustness, I also proceed by both reducing and increasing the number of instruments used (in sequential stages). Bond suggested that if there is a clear pattern in what happens to each coefficient (e.g. monotically increasing or decreasing as more instruments are used) the mentioned bias maybe relevant. Such pattern was not found. For simplicity, the iteration results are not reported in the paper.

[18] An orthogonal deviation x_{it}^{*} is given by:

$$x_{it}^{*} = \left[x_{it} - \left(x_{it+1} + ... + x_{iT}\right)/(T-t)\right](T-t)^{\frac{1}{2}}/(T-t+1)^{\frac{1}{2}} \qquad t = 1, ..., T-1$$

An orthogonal deviation is the deviation of the observation from the average of future observations in the sample, $x_{it} - \frac{(x_{it+1}+....+x_{iT})}{T-t}$; this deviation is then weighted (to standarize the variance) by multiplying it by $\left(\frac{T-t}{T-t+1}\right)^{\frac{1}{2}}$. If the original errors are IID, so will be the errors using orthogonal deviations. See, Arellano and Bover (1995), for further details.

Groups is likely to be biased downwards if the AR(1) provides a good representation for the series. However, the differenced GMM estimator is found to be significantly below the lower bound indicated by the Within estimator. These downward biases in differenced GMM estimates of the AR(1) are consistent with the finite sample biases expected in the case of highly persistent series. The preferred specification is clearly the GMM system estimator. With the introduction of the equations in levels I obtain a remarkable improvement in the precision of the parameter estimates. The results indicate a large degree of persistence in the net interest margins (.728-.759). Such inertia may arise from lagged effects of the explanatory variables, normal in balance sheets data with annual frequency. We can also observe that the assumption of disturbances serially uncorrelated cannot be rejected. As this model is overidentified, we can use the Sargan statistic to test the validity of the overidentifying restrictions. In this case, I obtain a chi-square statistic, giving the reported p-value of 0.123. Consistent with the serial correlation tests, the null hypothesis that these moment conditions are valid is not rejected at any conventional level.

Basic Model A simple approach, and first step, to study the pattern of the margins through the business cycle fluctuations is to include GDP growth in the AR(1) model. Prior to presenting the results, I would like to clarify the interpretation. To the extent that the assumptions regarding the instruments employed are correct; the econometric methodology is designed to isolate the effect of the exogenous component of the explanatory variable on the interest margins. Hereafter, when I mention the "impact" or "effect" of a given variable on the margins, I am referring to this isolated exogenous component and not merely describing the association between both. Table 2 present the results. *GROWTH* has a highly significant and negative effect on the margins. According to the preferred system specification, an increase in income by 10% causes the margin to fall by approximately 1% (100 bp) on impact.

The literature shows extensive evidence strongly linking long-run economic growth and financial development. We may be worried that these intermediation improvements may explain the negative relationship mentioned above. In order to asses and control for financial development, I will proceed by including a *three-year-overlapping* average of private credit offered by commercial banks (*PRIV.CRED (avg))* in the conditional set.[19] The results are depicted in Table 3. Notice that by computing averages to this dependent variable we are artificially constructing a persistent series aimed not to be affected by short-run fluctuations. Finite sample bias are therefore expected in the differenced specification. Contrary, the preferred system specification characterizes this variable to have a significant and sizeable negative effect on the margins.[20] Tough, with a slightly lower coefficient, *GROWTH* remains significantly at a 10% level. I explain this result supporting the hypothesis of fluctuating margins at a business cycle frequency.

Sensitivity analysis To asses the strength of the countercyclical nature of the margins depicted in the basic model I will use various conditioning information sets. I will start by introducing a proxy for concentration as a control variable. It measures the assets of the three largest banks as a share of assets of all commercial banks in the system. The results are in Table 4. The variable is significant and again enters with a negative sign. These results support the implications of the E-S hypothesis which predict a variety of efficiency gains from banking consolidation. The large number of explanatory variables accompanied with a relative large p-value for Sargan estimates raises a concern for overfitting bias. However no clear pattern in what happens to each coefficient is observed when reducing or increasing the number of instruments. For additional robustness, I report results for an equivalent system estimator based in a combination of orthogonal deviations and level equations. No significant changes are obtained.

If the negative effect of concentration is explained by efficiency gains. I would expect this impact to vanish when controlling for operational costs. In table 5, I expand the conditioning set and include Overhead Costs *(OVER-COSTS).* They measure personnel expenses (mostly wages) and other non interest expenses over total bank assets in the local bank system.

As expected large operational costs causes margins to increase. The variable enters significatively and with a sizeable coefficient. Providing support to the E-S hypothesis, the inclusion of this variable breaks down the independent contribution of concentration turning it insignificant and small in sign.

[19] *Three-years-overlapping* averages are calculated as: $x_{it(avg)} = \frac{x_{it}+x_{i,t-1}+x_{i,t-2}}{3}$.

[20] The coefficient can be interpreted as follows: If the *Priv.Credit/GDP* ratio increases by 50%, margins fall by approximately 150 basis points on impact.

Although the size of *PRIV.CRED (avg)* is marginally reduced, the cyclical component (*i.e. GROWTH*) remains mostly unaffected and significant at a 5% level.

Saunders and Shumacher (2000) show that interest rate volatility, usually observed in a context of high and variable inflation is positively related with margins. Thus, I proceed by adding inflation and real interest rates to the conditioning set. Nonetheless, none of these variables turns out to be significant. See again Table 5.[21]

The Role of Foreign Entry Up to this point, I have shown that the exogenous cyclical component of economic growth is negatively associated with net margins. Moreover, the mentioned link is not due to potential biases induced by omitted variables (including that derived from unobserved country specific effects); simultaneity or reversed causation. In the next step, I will test the main hypothesis proposed in the paper. That is, countercyclical markups are the result of a limit pricing aimed to deter entry of competitors in a segmented local financial system. As I explained in the introduction, although the threat of entry is a non-measurable concept, foreign penetration can be considered a good proxy for it. Consequently, I would expect the negatively association found up to this point to vanish when controlling for foreign entry.

Thus, I move forward and introduce foreign entry in the conditioning set. See tables 6A-6B. Specifically, the covariate $ForeignBanks$ refers to the number of foreign banks divided by the total number of banks in a given country. Foreign bank entry is measured as a *change* in foreign bank presence (i.e. $\Delta ForeignBanks_{it}$). The first experiment, not reported here, consisted in introducing $\Delta ForeignBanks_{it}$ into the extended model presented in the last subsection. It turns out that such covariate exerted negligible and non-significant effect. Different are the results when, instead, I consider $\Delta ForeignBanks_{i,t-1}$. See the last three columns of Tables 6-A and 6-B. Therefore, if limit pricing exists, it occurs one year after entry decision is ef-

[21] It is puzzling to observe that the coefficients for *GROWTH* and *PRIV.CRED (avg)* actually increase when these covariates are included to the conditioning set. There is no a clear justification. It might be the result of money-based disinflation programs being accompanied by short-lived recessions. These events would imply, at the same time, higher margins due to the recession but lower margins and credit availability resulting from stable and low inflation. Thus, if we do not control for inflation, we would expect margins to be less countercyclical and less sensible to variables linked with growth indicators.

fectively taken. Such covariate not only exerts a significative negative effect on the margins, but also breaks down the independent impact of *GROWTH* by turning it small and non-significant. It maybe the case that the beginning of wholesale operations occurs some time after the official entry registration occurs. These results may also provide support to the supposition of entry occurring at a wholesale level but spreading in retail niches with a time lag.

To asses the strength of this last critical finding, I would like to discard the possibility of lower margins a consequence not of *entry* but the solely result of pro-efficiency gains from a larger *presence* of foreign banks in the local financial structure.

Notice that the proposed model here is:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \gamma \Delta ForeignBanks_{i,t-1} + (\eta_i + \varepsilon_{it}) \quad i = 1, 2, ..., N;\ t = 2, 3, ..., T. \tag{5}$$

where y_{it} is the dependent variable, x_{it} any of the controlling sets already introduced, and $\Delta ForeignBanks_{i,t-1} = ForeignBanks_{i,t-1} - ForeignBanks_{i,t-2}$.

Alternatively, (5) can be expressed as:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \gamma ForeignBanks_{i,t-1} - \gamma ForeignBanks_{i,t-2} + (\eta_i + \varepsilon_{it}). \tag{6}$$

This setup is presented in the first columns of Tables 6-A and 6-B. Thus, if the results are driven by *entry* we expect the coefficients, γ, preceding $ForeignBanks_{i,t-1}$ and $ForeignBanks_{i,t-2}$ being both significant, of the same magnitude but with opposite signs (i.e. the first one negative and the second one positive). Contrary, if the results are driven by the *presence,* $ForeignBanks_{i,t-2}$ must be either negative or at least non-significant.

Once again, the results do not reject the validity of my hypothesis. Pro-competitive effect of entry in the local banking system is short lasting and immediately vanishes after one year.

Regional Analysis To asses the robustness of the results shown in the previous subsections, I test whether the cyclical pattern of the margin and the effect of entry differ across different groups countries. I just make the distinction between developing countries and developed ones. This simplification is aimed to restrict the number of covariates, reduce the number of instruments and avoid any risk of overfitting bias. I adopt the convention found in the World Development Indicators 2002 which divides all the displayed countries in four different income groups. Here, the first quarter will be regarded as "developed countries" while all the others will belong to the group of "developing countries". Dummies will be included. Specifically, $POOR$ will be equal to one if the country belongs to the latter group and zero otherwise. In what follows, the estimated model will be:

$$y_{it} = \delta_1 \left[z_{it} * POOR \right] + \delta_2 \left[z_{it} * (1 - POOR) \right] + (\eta_i + \varepsilon_{it}). \qquad (7)$$

where z_{it} is just the vector including the lag dependent variable plus any conditioning set of current or lagged covariates; and δ_1, δ_2 the collection of estimated parameters for developing and developed countries respectively.[22]

In table 7-A, I present the results for the basic model at a regional level. Again, the persistency of the dependent variable generates finite sample biases in the differenced estimator. However, in both specifications, the negative effect of *GROWTH* in margins is significantly larger for developing countries. In table 7-B, I control for financial development and, subsequently, I add concentration as previously done in the extended model. For robustness, I also report orthogonal deviations. The results indicate we can not reject the hypothesis that the coefficients for both group of countries are not significantly different. The only minor exception are the biased *two-steps* estimates of *GROWTH* for developed countries. Since this group of countries is relatively small the usual asymptotic distribution approximation seems no to be reliable in this case.

The next step is to consider the effect of entry. Since the estimates of the extended model do no significantly differ, I group them again (so as to reduce the number of explanatory variables and gain efficiency), and look for

[22] Actually, arranging terms we can interpret (7) as follows; if the country is developing, $POOR = 1$ and $y_{it} = \delta_1 z_{it} + (\eta_i + \varepsilon_{it})$; if the country is developed $POOR = 0$; so that:$y_{it} = \delta_2 z_{it} + (\eta_i + \varepsilon_{it})$.

each groups' reaction to foreign bank penetration. The results are in Table 8. Again the introduction of foreign entry breaks down the effect of growth. Nonetheless, entry significantly lower margins only in developing countries. Pro-efficiency gains from entry are not observed in developed countries. For this group of countries, the coefficient associated with entry is low, only marginally significant and positive. In different words, if my hypothesis is valid, it is only the result of entry in developing countries.

These results are in coincidence with the predictions of Claessens et al (2001). They argue that developed country banking markets tend to be more competitive with more sophisticated participants. If there is any technical advantage, foreign banks posses at the time of entry, they are not significant enough to overcome the information disadvantages they face relative to domestic banks. In developing countries, pervasive market inefficiencies and outmoded banking practices allow foreign banks to outweigh such information disadvantages.

Put it in a different perspective. In the last group of countries, the reach of financial development is severely restricted, the banking systems are small as well as likely to be subject to a more bothersome regulation. If foreign banks takes the decision to participate in these local markets, despite all these impediments, the resulting threat of competition for retail niches is significantly larger.

However, one puzzling result arises in this analysis. Why do we still observe a moderate countercyclicality margins' pattern in developed countries? One possible explanation is that these markets are more efficient, competitive and contestable. If a boom occurs, the financial customer base expands and the system fully exploits efficiency gains derived from economies of scale and scope. Such gains are translated in lower margins. In any case, this topic deserves further research.

One last question remains to be addressed in the empirical analysis. All the conclusions I outlined rely on the assumption that entry of foreign banks occurs in booming periods. Since the effect of entry is sizeable only in developing countries, I will restrict the sample just to include all of them. The results are in table 9.

In the first three columns we can observe that, even after controlling for institutional variables, the state of the economy significantly affects the entry decision. That is, *GROWTH* remains significant at a 5 % level. Again, we rule out this result being drive by simultaneity or reversed causation. Under the assumption that institutional variables do not significantly vary in a the short span of time existent in the panel, we can include non-time varying covariates in the regressions. For robustness, I also include financial development, market concentration and the GDP *level* in the conditional setting. The last control is aimed to capture the size of the local financial sector. Again, I do not have evidence to reject the assessment that the current state of the economy triggers entry decision. From the results, we can also read that institutional factors do not play a significant role in the amount of foreign entry. However, it seems that the size of the economy plays a role. I interpret that the larger the economy, the larger the size of the financial market and, thus, the bigger the space available for additional competitors.[23]

6 A simple theoretical model

In this section I will present a simple general equilibrium model aimed to reflect the comprehensive role of the proposed "bank lending channel" in the economy. I will start from a standard DSGE Real Business Cycle model with variable labor supply in the spirit of Hansen (1985). Then, I will move forward and introduce imperfect competition with limit pricing in the financial system. This modification will create a disintermediation between borrowers and entrepreneurs that will amplify the response of the real variables to technology shocks.

6.1 Households

The household sector is conventional. There is a continuum of households of unit length. Each household works, consumes, and invests its savings in regular deposits.

[23] A final caveat. Since entry is procyclical and therefore co-linear with growth, someone could argue that the inclusion of entry (as a control variable) would necessarily break down the effect of growth in the margins. In that circumstance, countercyclicality would not be the result of limit pricing but the effect of omitted variables not included in the conditioning sets. Altough, I carefully controled for all known reasons that could generate a countercyclical pattern (so as to discard this criticism); I admit that only dissagregated data can partially disentangle and put in evidence limit pricing strategies. I postpone this issue for future research.

The representative household maximizes:

$$E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{1-\gamma} C_t^{1-\gamma} - \frac{a_n}{1+\gamma_n} N_t^{1+\gamma_n} \right] \tag{8}$$

Subject to the budget constraint:

$$C_t + D_{t+1} = W_t N_t + R_t D_t + \Pi_t. \tag{9}$$

C_t is consumption; N_t is labor supply; W_t denotes the real wage; D_t are deposits (in real terms) held at financial intermediaries and R_t is the real interest rate paid to depositors. Π_t are real dividends payments from ownership of financial intermediaries.

6.1.1 Optimality Conditions

Household behavior obeys:

Consumption and saving intertemporal allocation;

$$1 = \beta E_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\gamma} R_{t+1} \right\}. \tag{10}$$

Labor allocation;

$$W_t C_t^{-\gamma} = a_n N_t^{\gamma_n}. \tag{11}$$

6.2 The Entrepreneurial Sector

Entrepreneurs construct capital in each period for use in the subsequent period. Capital is used in combination with labor to produce output. Entrepreneurs are risk neutral. Assuming CRS, Cobb-Douglas technology, the aggregate production function is:

$$Y_t = A_t K_{t-1}^{\alpha} N_t^{1-\alpha}, \tag{12}$$

Where Y_t is aggregate output, K_{t-1} is the aggregate amount of capital constructed by entrepreneurs in period $t-1$, N_t is the labor input, and A_t is an exogenous technology shock.

Thus labor demand satisfies,

$$(1-\alpha)\frac{Y_t}{N_t} = W_t. \tag{13}$$

Demand Curve for New Capital The construction of new capital is determined by the level of investment I_t. Thus, the capital stock obeys:

$$K_t = I_t + (1-\delta)K_{t-1}, \tag{14}$$

where δ is the depreciation rate.

The gross return to holding one additional unit of capital from t to $t+1$ can be written as:

$$R_{t+1}^k = E_t\left\{\frac{\alpha Y_{t+1}}{K_t} + (1-\delta)\right\}. \tag{15}$$

Supply Curve for New Capital The entrepreneur's demand for capital satisfies the following optimality condition:

$$R_{t+1}^{k} = (1 + \Xi_{t+1}) R_{t+1}, \tag{16}$$

where the real interest rate, R_{t+1}, is the gross cost of funds absent imperfect competition in the financial system and $(1 + \Xi_{t+1})$ is the gross markup charged by the intermediary bank. I assume that new equity and bond issues are prohibitively expensive, or not available for local firms, so that all investment finance is done with bank credit. I will ignore the presence of the bank multiplier and the existence of reserves: The overall amount of credit in the economy must be equal to the overall amount of new household deposits, i.e.

$$D_{t+1} = I_t. \tag{17}$$

I assume that the banking system is highly segmented in a large number of niches, n, each one served by an established bank with monopoly power. The optimal net markup set by the representative incumbent bank , Ξ_{t+1}, is the result of a limit pricing strategy aimed to deter entry and satisfies:

$$\Xi_{t+1}\left(\frac{I_t}{n}\right) - \lambda\left(\frac{I_t}{n}\right)^{1-\tau} = 0 \qquad\qquad 0 < \tau < 1. \tag{18}$$

where λ represents the cost of serving a niche for banks that are outside of the niche (working at a wholesale level), and τ reflect economies of scope and scale in the banking industry. I postpone the microfoundations of (18) to the next subsection.

Resources Constraints The resource constraint of the economy is:

$$Y_t = C_t + I_t. \tag{19}$$

6.3 The Financial System

As already mentioned, I assume that the banking system is highly segmented in a large number, n, of sectors or regions (niches).

The size of each niche is the same, and each of them is served by an established bank (incumbent), i, that possesses a local monopoly and therefore finances an equal fraction $\frac{I_t}{n}$ of the total investment. By assumption, each incumbent can only serve its own niche.

This intermediary chooses a net markup for its niche, Ξ_{t+1}, at the beginning of period t and obtains a revenue $\Xi_{t+1}\left(\frac{I_t}{n}\right)$. I assume that the cost of serving its own niche for each bank i is

$$\upsilon_i\left(\frac{I_t}{n}\right)^{1-\tau}. \tag{20}$$

The bank system possesses operational economies of scope and scale. Thus, I assume that $0<\tau<1$. On the other hand, υ_i captures any idiosyncratic operational (in)efficiency and information (dis)advantages any bank may have. The *cost-efficiency level*, υ_i, is private information and is not revealed to banks outside the niche.

Entry and mergers I assume that entry is possible in this banking system. However, it occurs in successive stages. Entrants in the *banking system* at time t only start competing in the *niche* at time $t+1$, which introduces a one-period time-to-build lag in the model. However, right after the entry decision is effectively taken, the *just arrived* is already inside the banking system (but only at a "wholesale level"). Hence, during period t it is able to temporarily serve *any* of the n niches until finally established in one of them in $t+1$. The aim is to capture the idea of entry taking place in the wholesale market first and spreading later to retail segments (niches).

The entry stages occur as follows:

A) At the beginning of period t, a potential competitor, j, attempts to enter in the banking system. At no cost, it draws its *cost-efficiency level*, υ_j, from a common uniform distribution $U(\upsilon)$ with support on $[0,\lambda]$.

B) After learning its own υ_j, the potential competitor chooses whether to enter and "fight" for one of the niches or withdraw from the banking system.

For simplicity, I assume that the entrant is able to enter only one of the niches (i.e. multi-sectorial entry is not possible). The closer υ_j is to zero, the more efficient the potential entrant would be, and the easier to take-over a niche. I will assume that the number of total draws in the banking system will be large enough so that *at least* one potential competitor effectively takes the decision to fight every period.

C) I will assume that an outsider is able to enter any of the niches at $t+1$, after incurring in fixed sunk entry costs, sk_t, in period t.[24] sk_t is exogenous and is in output units. We can also interpret changes in sk_t as changes in entry regulations.

Notice that incumbents' $\upsilon_i s$ are unknown to the *just arrived*. Hence, entrants are indifferent when choosing the particular niche to "fight" for. Thus, I assume that once inside the banking system they randomly choose which particular niche to enter at the end of period t.

D) During period t entrants are able to temporarily serve in *any* of the n niches until finally established in one of them. The general cost of serving other niches is:

$$\lambda \left(\frac{I_t}{n} \right)^{1-\tau}, \qquad (21)$$

where $\lambda \geq \upsilon_i$ for every i, in accordance with the assumption introduced in (A). As in Petersen and Rajan (1995), I assume that banks that are physically closer to the customers have lower costs of monitoring and transacting with both firms and depositors.

E) At the very beginning of period $t+1$, the entrant is inside the niche an is able to "learn" the incumbent's υ_i. Bertrand competition occurs straightaway. The following proposition yields the main result for this case.

[24] As I said, we can include in them advertisement costs aimed to gain reputation or the construction of a network of branches and ATMs.

PROPOSITION 1 *Under Bertrand competition, only two possible outcomes are possible. If $\upsilon_j > \upsilon_i$, the entrant fails and is forced to merge. If $\upsilon_j < \upsilon_i$ the entrant successfully displaces the incumbent and forces it to merge. The optimal strategy for the looser is to merge immediately and not to compete. The only visible outcome is the possible change of the incumbent at the very beginning of $t+1$.*[25]

Proof. See Appendix 1.

E) The successful new incumbent keeps the niche until it is hit by an exit-inducing shock that occurs with probability $\delta_D \in (0,1)$ in every period.[26] For simplicity, I do not model endogenous exit that is not driven by the mentioned Bertrand competition. The "death" shock is independent of the bank's efficiency level. I assume that the empty niche left by every "dead" firm is immediately filled by an entrant. Right after drawing an efficiency level, the entrant is able to use the existent network left by the "dead" firm (avoiding any sunk costs as well as the time-to-build lag). The number of banks and the frequency of "death" is high enough so that $E(\upsilon_i) = \frac{\lambda}{2}$, and $U(\upsilon)$ nests the cost-efficiency distribution of all incumbents in the financial system.

Limit Pricing The outstanding feature of this setup is that entry and pricing decisions are independent. The potential competitor knows the *cost-efficiency* level distribution $U(\upsilon)$ of the banking system, but not the particular $\upsilon_i s$ of each incumbent. In this circumstance, the optimal pricing strategy, Ξ_{t+1}, for the incumbent is to set every period a markup low enough to make sure that any of the *just arrived* will not obtain any expected positive profits if it decides to serve the whole niche offering a net markup below Ξ_{t+1}. That is:

$$\Xi_{t+1}\left(\frac{I_t}{n}\right) - \lambda\left(\frac{I_t}{n}\right)^{1-\tau} = 0. \quad (22)$$

Therefore, pricing decision is exactly the same in all the niches. Since by assumption, all the niches are of the same size, we can interpret this relationship as the pricing decision taken by the representative bank of this

[25] By Definition the point likelihood of $\upsilon_j = \upsilon_i$ is null.

[26] This endogenous entry-exogenous exit mechanism is taken from Ghironi and Melitz (2004).

economy. This expression coincides with (18) in the general equilibrium setup.

Hence, for every period, t, the expected profits for each incumbent i are:

$$\pi_{i,t} = (\lambda - \upsilon_i) \left(\frac{I_t}{n} \right)^{1-\tau} > 0. \qquad (23)$$

Equations (22) and (23) can be interpreted as follows: the greater aggregate investment, the bigger the size of the niche, and the higher the competitive pressure of the *just arrived*. In turn, this forces the incumbent to offer lower markups. However, the level of profits, $\pi_{i,t}$, is independent of the markup levels and is directly related to the size of aggregate investment.[27]

Entry decision Banks are forward looking and correctly anticipate their stream of profits. After drawing a υ_j, a potential entrant will take the decision to fight if and only if the expected post-entry present discounted net value of the expected stream of profits $\{\pi_{j,t}\}_{t=1}^{\infty}$ is positive:

$$V_{j,t} = \left\{ E_t \sum_{t=1}^{\infty} \left[\beta \left(1 - \delta_D\right)\right]^t \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \pi_{j,t} \right\} \left(1 - \frac{\upsilon_j}{\lambda}\right) - sk_t > 0. \qquad (24)$$

Banks discount future profits using the household's stochastic discount factor, adjusted for the probability of survival. The pre-entry probability of "defeating" the incumbent and taking-over the niche is $1 - \frac{\upsilon_j}{\lambda} = \Pr(\upsilon_j < E(\upsilon_i))$. Equations (24), and (23) imply that entry is procyclical (i.e. entry increases when the level of investment is high). The larger the discount factor and the probability of the exit-inducing shock the stronger the procyclicality.

Entry is affected by market regulation that alters the value of sk_t. Equation (24) implies that the higher sk_t, the lower the amount of entries in the banking system (and vice versa). But the higher sk_t, the more likely these entries are successful when fighting for the niche.

[27] For the representative bank of the economy, the level of profits is given by $\pi_t = \frac{\lambda}{2} \left(\frac{I_t}{n} \right)^{1-\tau} > 0$.

By setting $sk_t \to \infty$, the government effectively prohibits entry in the banking system. In this case, countercyclical limit pricing is not necessary, and incumbent banks are be able to practice standard short-run profit maximization every period. However, to keep it simple I do not analyze this case.

6.4 Model Parametrization

The only distinctive aspect of the general equilibrium model relative to a benchmark RBC setup is the limit pricing scheme in the financial system, characterized by equations (16) and (18). The former characterizes how imperfect competition in the financial system influences capital demand. The latter describes the limit pricing strategy chosen by the incumbent banks. If we restrict the net financial markup Ξ_{t+1} to zero in equation (16), we effectively shut off the "bank lending channel" and the model reverts to a conventional RBC model.

I set the quarterly discount factor β to 0.99 (which also pins down the steady state quarterly real interest rate depositors receive $R = \beta^{-1}$). Average hours worked relative to total hours available are set equal to $\frac{1}{3}$. I set the the elasticity of intertemporal-substitution, $\frac{1}{\gamma}$, equal to one, and γ_n equal to zero. Following Hansen (1985), I set the standard deviation of the productivity innovations to 0.712. The capital share, α, is 0.36. The quarterly depreciation, δ, is assigned the value of 0.025. Finally, following the descriptive statistics for developing countries, I set the quarterly steady state net financial markup equal to 142 basis points and choose $\tau = 0.70$.

6.5 A Negative Technology Shock

I consider an unanticipated one percent decrease in technology. I assume further that the shock obeys a first order auto-correlation process that persists at the rate of 0.95 per quarter. In Figures 1-2, I plot the response of the eight endogenous variables under both perfect and imperfect competition in the financial system. As I said, the former exactly resembles the basic RBC specification. In this case, financial markups are null and the "natural" or "wicksellian" interest rate depositors and entrepreneurs face do not differ.

In the competitive model, a negative technology reduces output, factor productivity and consumption today by more than in future periods. Output and consumption therefore fall today and return later to their original levels. Households would like to smooth their consumption and attempt to shift resources away from future periods to current period. For this reason, we would expect the natural real interest rate to be pushed up.

But, on the other hand, investment demand goes down because the technology shock has decreased the outcome of production. By itself, this pushes down the natural interest rate; offsetting the pressure that comes from households' desire to substitute consumption away from future periods. The net effect of these counteracting pressures is to slightly decrease the natural interest rate in just 7 basis points.

With imperfect competition the results change. The monopolistic intermediary has the possibility of providing credit after charging a markup over the interest rate paid to depositors. The intermediary banks allow households to substitute consumption away from other periods toward this period by decreasing significantly the interest rate paid on deposits. As a result, consumption does not initially fall as much as in the competitive model. But this relatively higher consumption lowers the marginal utility of income and reduces the work effort even more.

A decrease in the labor input negatively affects the outcome of production and the productivity of capital. This is the cause for an even lower demand for investment relative to the baseline case. Under perfect competition, a resulting lower investment demand and lower interest rates paid to depositors would be reflected in a sharp decrease in the interest rate entrepreneurs face. However, the fact that investment falls and the financial market shrinks causes the threat of entry to decline, and higher markups are compatible with the limit pricing scheme. The financial markups increases 9.42% (13 bp) on impact. The higher markups do not allow costs of borrowing for entrepreneurs to fall significantly, and throughout the experiment, the optimal capital stock is smaller than in the competitive case and the volatility of all real variables is higher.

6.6 Volatility and Welfare

Macroeconomic Variability and Sensitivity Analysis Quantitative results presented in tables 10 and 11 confirm that the presence of monopoly power and countercyclical markups in the intermediary sector ends up increasing the volatility of all real variables amplifying the business cycle relative to the simple RBC model.

In my parametrization, the role of τ is critical for the countercyclical nature of the markups. The larger τ, the larger the banking economies of scale and the higher (lower) the possibility of outsiders working at an efficient scale in a booming (recessionary) economy. In turn, this causes the incumbent to set relatively lower (higher) markups. As expected, in table 12 we can observe how the volatility of real variables monotically increases when τ increases.

Welfare Results Now I consider how the welfare of the representative household is affected by the presence of monopolistic power in the banking system. I solve the model using a second-order approximation as in Collard and Juilliard (2001). Otherwise, conventional linearization can generate approximation errors that may be the cause of possible welfare reversals (see Kim and Kim, 2003 for details). The welfare criteria considered here is based on a second-order Taylor expansion of the representative household's expected utility function (8), $W_t = EU_t$, around the deterministic steady state values.

$$\begin{aligned} W_t &= \frac{1}{1-\gamma}\bar{C}^{1-\gamma} - \frac{a_n}{1+\gamma_n}\bar{N}^{1+\gamma_n} + \bar{C}^{1-\gamma}E(\hat{c}_t) \\ &\quad -a_n\bar{N}^{1+\gamma_n}E(\hat{n}_t) - \frac{1}{2}\gamma\bar{C}E(\hat{c}_t^2) - \frac{1}{2}\gamma_n\bar{N}^{1+\gamma_n}E(\hat{n}_t^2). \end{aligned} \tag{25}$$

Where $\bar{C}$ and $\bar{N}$ are steady state values of consumption and labor and hats denote percentage deviations from the steady state. When evaluating the welfare criteria I get -0.084 and -0.039 for the original parametrization under imperfect and perfect competition respectively. These negative values reflect just the ordinal nature of the utility function that can easily be reversed after a standard monotonic transformation. In any case the results confirm that the representative household is better off with perfect competition. A monopolistic environment affects welfare of the household trough two different channels. Firstly, the financial markup generates a permanent disintermediation between borrowers and entrepreneurs that results in lower steady state levels of capital accumulation, output and hence consumption.

Secondly, the countercyclical pattern of such markups increases the volatility of real variables and thus reduces welfare.

7 Conclusions

The contestability of the retail banking sector is obstructed for the requirement of incurring in large sunk entry costs in highly segmented markets. In turn, this implies that right after entering, the banks must capture a significant fraction of the market to make the branch or ATM net workable. The idea of this paper is that limit pricing strategies aimed to avoid competition in banking retail niches are adopted when incumbents face any entry threat. During recessions the actors in the local banking system are more able to exert their monopolistic power; but booming periods led to an expansion of the financial system that allows entrants to work at an efficient scale. Contestable markets force incumbents to lower markups so as to deter entry. In turns, this generates countercyclical financial markups.

Using annual aggregate bank data for a large set of countries for the period 1990-2001; I find that financial markups are strongly countercyclical even after controlling for simultaneity, financial development, banking concentration, operational costs and inflation. Since threat of entry is a not a mensurable concept, I use foreign penetration as a proxy. Here, I exploit the evidence that foreign bank entry initially takes place in the wholesale market and expands later to some of the retail niches. Effectively, in the empirical models I find that entry (and not presence) of foreign banks is the omitted variable that disentangle the cyclicality of the markups. In a regional analysis, I find that the efficiency gains resulting from foreign entry are only significant in developing countries where banking systems are usually small, riskier, and subject to a bothersome regulation. Thus, I interpret that if foreign banks takes the decision to participate in these local markets, despite all these impediments, the resulting threat of competition for retail niches is significantly larger.

The modelling of the financial system, captures several of the features of the empirical evidence. In the theoretical model, entry occurs at wholesale level and then spreads to the retail system. The market is highly segmented in niches and more efficient entrants end up taking-over current incumbents. Entry is procyclical and more likely to occur in deregulated markets, but actually, is more effective and successful if markets are regulated. Although, changes in the market structure do not affect the markups, the threat of entry forces incumbents to set prices that deter entry. Finally, economies of scale facilitate entry in booming periods, and vice versa, generating countercyclical markups. At a general equilibrium level, the behavior of this financial system generates a "bank lending channel" that increases the volatility of real variables, amplifies the business cycle and reduces welfare. Credit is more expensive during recessions, and firms and households postpone investment and work decisions pushing the recession further.

Not counting with bank-level disaggregated data was a considerable constraint for this study. For instance, it would be interesting to study whether the regional markets that are more concentrated, with low degree of financial development, or more regulated reflect different cyclical patterns. In a different perspective, it is worth to remark that efficiency gains from financial liberalization and market de-segmentation may be offset by some important negative effects not considered in this study. For instance, regional banks are engaged in long term relationships with small domestic entrepreneurs which otherwise would be out of the credit market. It follows that entry threats that force low margins can increase the degree of banking fragility and disrupt these relationships. I postpone these issues to future research.

8 Appendix 1

Proof of Proposition 1

Define the break-even level of margins θ_i and θ_j for the incumbent and the entrant. The break-even level is equal to the value of the net margin that provides them zero profits when serving all the niche. That is:

$$\theta_i \left(\frac{I_t}{n}\right) - \upsilon_i \left(\frac{I_t}{n}\right)^{1-\tau} = 0, \text{ and} \qquad \theta_j \left(\frac{I_t}{n}\right) - \upsilon_j \left(\frac{I_t}{n}\right)^{1-\tau} = 0. \qquad \text{(A.1)}$$

Now, let's analyze the case where $\upsilon_j > \upsilon_i$, and thus $\theta_j > \theta_i$.

Consider for example, $\Xi^i_{t+1} > \Xi^j_{t+1} > \theta_j$. The bank i has no demand and its profits are zero. On the other hand, if bank i charges $\Xi^i_{t+1} = \Xi^j_{t+1} - \varepsilon$ (where ε is positive but nil), it gets the entire niche and has a positive profit $\Xi^j_{t+1} - \varepsilon - \theta_i > 0$.

Therefore bank j cannot be acting in its own interest by charging Ξ^j_{t+1}. Now suppose $\Xi^i_{t+1} = \Xi^j_{t+1} > \theta_j$, in that case they share the niche, and each one serves half of it. But, if bank j reduces its price slightly to $\Xi^j_{t+1} - \varepsilon$, it gets all the niche. Nonetheless, bank j will never charge $\Xi^j_{t+1} < \theta_j$, because it would make a negative profit. It follows that bank i can charge $\Xi^i_{t+1} = \theta_j - \varepsilon$ and guarantee for itself all the niche while obtaining a positive profit $\theta_j - \varepsilon - \theta_i > 0$.

Therefore bank j is indifferent where to stay or leave the niche, since will not be able to serve it. If Bank i offers Bank j a nil but positive amount ε of output so as to merge, it is in the best interest of bank j to accept it. Finally, a symmetric analysis holds when $\upsilon_j < \upsilon_i$.■

9 References

[1] Anderson, T. and Hsiao, C. "Estimation of Dynamic Models with Error Components" *Journal of the American Statistical Association*, 76, pp 598-606. 1981.

[2] Angelini, P. and Cetorelli, N. "Bank Competition and Regulatory Reform: The case of the Italian Banking Industry" *Banca d' Italia* WP 99-32. 2000.

[3] Angelini, P. and Cetorelli, N." The Effects of Regulatory Reform on Competition in the Banking Industry" *Journal of Money, Credit and Banking*, Vol. 35, No. 5. (Oct., 2003), pp. 663-684.

[4] Arellano, M. and Bond, S. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations" *Review of Economic Studies, 58, 277-297. 1991.*

[5] Arellano, M. and Bover, O. "Another Look at the Instrumental Variable Estimation of Error-Components Models" *Journal of Econometrics, 68, 29-52. 1995.*

[6] Bain, J. "A Note on Pricing in Monopoly and Oligopoly" *American Economic Review,* 39, 448-464. 1949.

[7] Bain J. "Barriers to New Competition". Cambridge, Massachussets: Harvard University Press, 1956.

[8] Bank for International Settlements. "The Banking Industry in the Emerging Market Economies: Competition, Consolidation, and Systematic Stability." BIS Papers 4, Basle. 2001.

[9] Barro, R. "Determinants of Economic Growth: A cross country empirical study" *The MIT press, 1997.*

[10] Baumol, W. "Contestable Markets: An Uprising in the Theory of Industry Structure ", *American Economic Review 72*(1), March, 1982. pp 1-15.

[11] Berger, A and Hannan, T."The Price-Concentration Relationship in Banking". *Review of Economics and Statistics.* 291-299. 1989.

[12] Bernanke, B and Blinder, A."Credit, Money, and Aggregate Demand" *American Economic Review,* May 1988, 78, 455-39.

[13] Bils, Mark. "Pricing in Customer Market" *Quarterly Journal of Economics,* November 1989, 104 (4), pp. 699-718.

[14] Blundell, R.; Bond, S. and Windmeijer,F. "Estimation in Dynamic Panel Data Models" Advances in Econometrics, V15, JAI Elsevier Science. 2000.

[15] Blundell, R. and Bond, S."Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" University College London Discussion Paper 97-07. 1997.

[16] Bond, S. "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice" *Portuguese Economic Journal. 2002.*

[17] Bowsher,C. "On Testing Overidentifying Restrictions in Dynamic Panel Data Models" Nuffield College Discussion Paper in Economics W28. 2000.

[18] Catena, M. "Banking Crises and Monetary Policy in Argentina", *mimeo,* UCLA. 1996.

[19] Cerasi, V.; Chizzolini, F. and Ivaldi, J. "Sunk Costs and Competitiveness of European Banks after Deregulation" Toulouse - GREMAQ Series Papers. 97. 473. 1997.

[20] Chevallier, J. and Scharfstein, D. "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence" *American Economic Review,* September 1996, 86 (4) pp 703-725.

[21] Claessens, S.; Demirguc-Kunt, A., Huizinga, H. "How does foreign entry affect domestic banking markets?" *Journal of Banking & Finance. 25.* 891-911. 2001.

[22] Demirguc-Kunt, D. and Huzinga, H. "Determinants of Interest Rate Margins and Profitability: Some International Evidence". Working Paper, Development Research Group, World Bank. 1998.

[23] Demirguc-Kunt, A.; Laeven, L. and Levine R."Regulations, Market Structure, Institutions, and the Cost of Financial Intermediation". *mimeo, 2003.*

[24] Demsetz, H. "Industry Structure, Market Rivalry, and Public Policy" *Journal of Law and Economics 16, 1-9.1973.*

[25] Dollar, D. and Kraay, A. "Trade, Growth and Poverty" Economic Growth Research. World Bank Group. 2001.

[26] Doornik, J.; Arellano,M. and Bond, S. "Panel Data estimation using DPD for Ox", *mimeo.* Nuffield College. December 2002.

[27] Edwards, S.Vegh, C. "Banks and Macroeconomic Disturbances Under Predetermined Exchange Rates" WP 5977. NBER. 1997.

[28] Focarelli, D.; Panetta, F. and Salleo, C. "Why do Banks Merge?", *mimeo,* Federal Reserve Bank of Chicago. May 1999.

[29] Focarelli, D. and Rossi, P. "La domanda di finanziamenti in Italia e nelle diverse aree del Paese (1984-1996)" Banca d'Italia, Temi di Discussione 333. 1998.

[30] Galeotti, M and Schiantarelli. "The Cyclicality of Markups in a Model with Adjustment Costs: Econometric Evidence for U.S. industry", *Oxford Bulletin of Economics and Statistics,* V 60 (2), 121-142. 1998.

[31] Gali, Jordi. "Monopolistic Competition, Business Cycles and the Composition of Aggregate Demand", *Journal of Economic Theory.* Volume 63(1). 73-96. 1994.

[32] Gertler, M. and Hubbard, R. "Financial Factors in Business Fluctuations" *mimeo,* Columbia -Graduate School of Business. 1988.

[33] Ghironi, F. Melitz, M. "International Trade and Macroeconomic Dynamics with Heterogeneous firms", *NBER Working Paper 10540. 2004.*

[34] Gilligan, T.; Smirlock, M. and Marshall, W. "Scale and Scope Economies in the Multiproduct Banking Firm" *Journal of Monetary Economics,* 13. 1984.

[35] Grady, D and Kyle, R. "The Simultaneity of Bank Decision-Making, Market Structure and Bank Performance". Journal of Finance 34, 1-18. 1979.

[36] Greenwald, B.; Stiglitz, J. and Weiss, A. "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations" *American Economic Review,* May 1984, 74(2), pp 194-99.

[37] Hancock, D. A. "Theory of Production for the Financial Firm." Kluwer Academic, Norwell, MA. 1991.

[38] Hannan, T. "Market Share Inequality, The Number of Competitors and the HHI: An Examination of Banking Pricing", *Review of Industrial Organization, 12, 23-35. 1997.*

[39] Hannan, T and Berger, A."The Rigidity of Prices: Evidence from the Banking Industry". *American Economic Review. 81. 938-945. 1991.*

[40] Hansen, G. "Indivisible Labor and the Business Cycle" *Journal of Monetary Economics*, 16, pp. 309-327. 1985.

[41] Holtz-Eakin, D.; Newey,W. and Rosen, H. "Estimating Vector Autoregressions with Panel Data" *Econometrica, 56, 1371-1396. 1988.*

[42] Jackson , W. "The Price-Concentration Relationship in Banking: A Comment". *Review of Economics and Statistics, 74, 1992. 371-376.*

[43] Keeley, M. and Zimmerman, G. "Competition for Money Market Deposit Accounts." *Federal Reserve Bank of San Francisco Economic Review* 1, 5-27. 1985.

[44] King, R. and Levine, R. "Finance and Growth: Schumpeter might be right" *Quarterly Journal of Economics 108,* 717-738. 1993.

[45] Levine, R.; Loayza, N. and Beck,T. "Financial Intermediation and Growth: Causality and causes" *Journal of Monetary Economics. 46* pp *31-77. 2000.*

[46] Murphy, K.;Shleifer,Andrei and Vishny, Robert."Building Blocks of Market Clearing Business Cycle Models". *Macroeconomics Annual,4,* pp *247-87. 1989.*

[47] Neumark, D and Sharpe, S. "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposit". *Quarterly Journal of Economics. 656-680. 1992.*

[48] Olivero, M. " Trade, Non-Competitive Banking and the International Transmission of Business Cycle", *mimeo,* Duke University. 2004.

[49] Petersen, M. and Rajan, R. "The Benefits of Lending Relationships: Evidence from Small Business Data" *Journal of Finance. 49. 3-37. 1994.*

[50] Phelps, E. and Winter, S. "Optimal Price Policy under Atomistic Competition" In E. Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory"* New York: W.W.Norton. pp. 309-37. 1970.

[51] Pigou, Arthur C. *Industrial fluctuations.* London: Macmillan and Co.1927.

[52] Prager, R. and Hannan, T. "Do Substantial Horizontal Mergers Generate Significant Price Effects? Evidence from the Banking Industry" *Journal of Industrial Economics 46, 433-452. 1999.*

[53] Ribon, S. and Yosha, O. "Financial Liberalization and Competition in Banking: An Empirical Investigation", Tel Aviv University, Working Paper. 23-99. 1999.

[54] Rhoades, S. "Structure- Performance Studies in Banking: A summary and Evaluation." Staff Economic Studies No 92, Federal Reserve Board, 1977.

[55] Rojas-Suarez, L. and Weisbrod, S. "Financial Market Fragilities in Latin America: from banking crisis resolution to current policy challenges" WP 94/117, IMF. 1994.

[56] Rotemberg, J. and Saloner,G. "A Supergame Theoretic Model of Business Cycles and Price Wars During Booms." *American Economic Review,* June 1986, 76(3). pp 390-407.

[57] Saunders, A. and Shumacher, L. "The Determinants of Bank Interest Rate Margins: an international Study" *Journal of International Money and Finance 19. pp 813-832. 2000.*

[58] Shaffer, S. " A test of Competition in Canadian Banking". *Journal of Money, Credit and Banking, 25, 49-61. 1993.*

[59] Simmons, K and Stavins, J. "Has Antitrust Policy in Banking Become Obsolete?" Federal Reserve Bank of Boston New England Economic Review, March, 13-26. 1998.

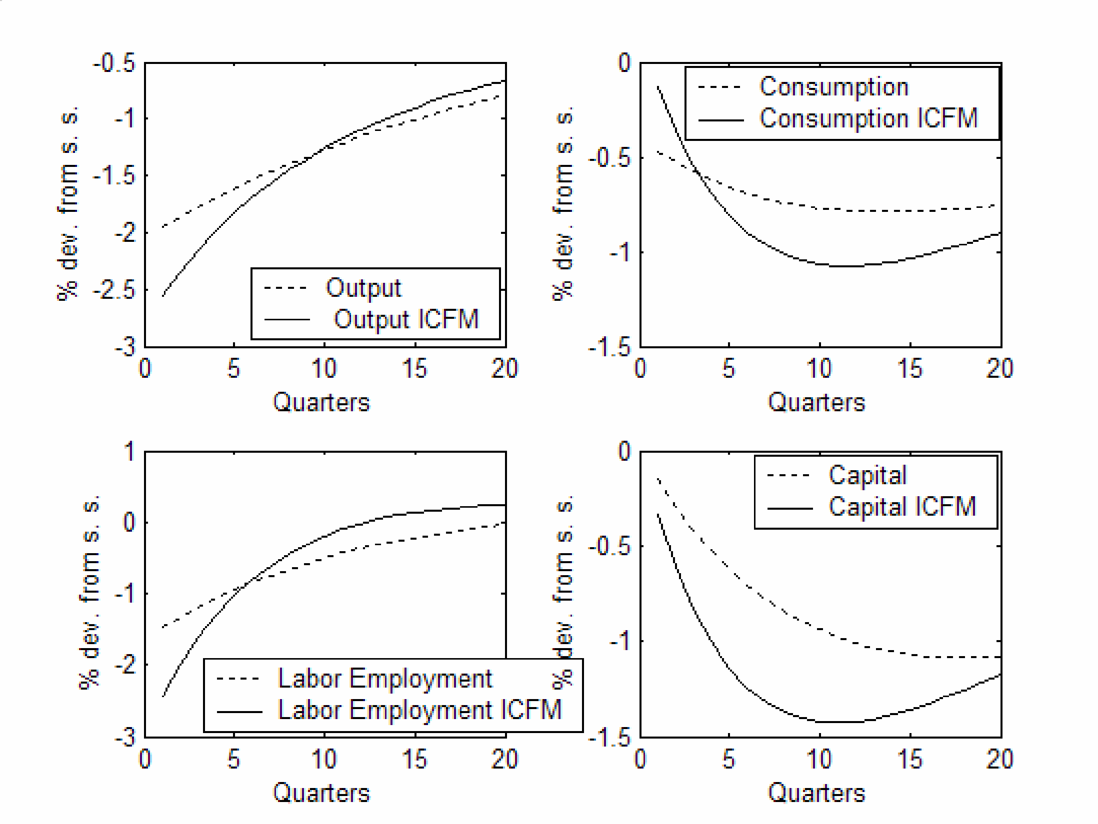
[60] Smirlock, M. "Evidence on the (Non) Relationship between Concentration and Profitability in Banking" *Journal of Money Credit and Banking* 17, 69-83. 1985.

[61] Spiller, P. and Favaro, E. "The Effects of Entry Regulation on Oligopolistic Interactions: The Uruguayan Banking Sector" *RAND Journal of Economics. 9. 305-327. 1984.*

[62] Vives, X. "Regulatory Reform in European Banking". *European Economic Review,* 35, pp 505-515. 1991.

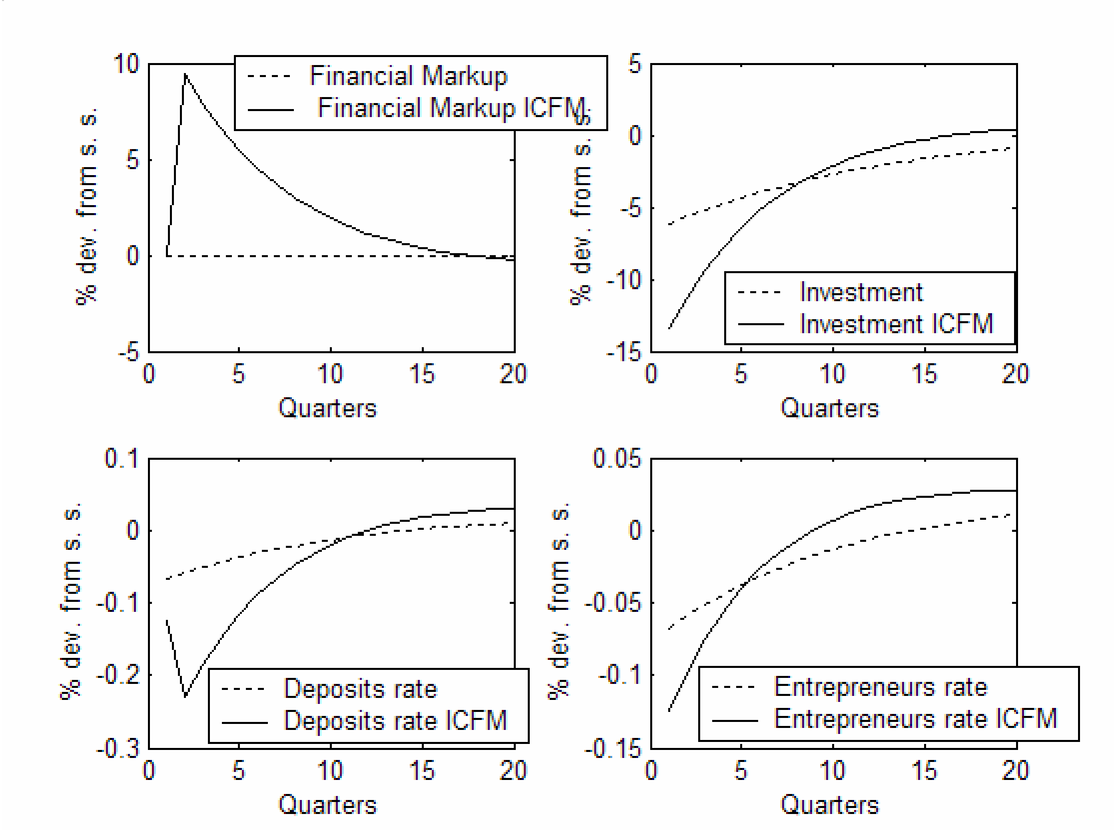
[63] Windmeijer, F. "A Finite Sample Correction for the Variance of Linear Two-Step GMM Estimators" Institute for Fiscal Studies. WP 00/19. 2000.

Figure 1



Percentage point response of the Monopolistic Financial Market (straight line), and RBC (dashed line) models' to an unanticipated 1 percent decrease in technology.

Figure 2



Percentage point response of the Monopolistic Financial Market (straight line), and RBC(dashed line) models' to an unanticipated 1 percent decrease in technology.

Regional Analysis
Descriptive statistics -Mean Values-

	NET INTEREST MARGINS	PRIVATE CREDIT	CONCENTRATION	No. of Foreign Banks	Obs.
Developing countries	**0.0571**	**0.2532**	**0.6645**	**0.3226**	**91**
Developed countries	**0.0268**	**0.7653**	**0.5751**	**0.2769**	**33**

Table 1: Time Series Properties.

Dependent Variable: Net Interest Margins (NIM_t)

	OLS LEVELS	WITHIN GROUPS	GMM DIF 1STEP	GMM DIF 2STEPS	GMM SYS 1STEP	GMM SYS 2STEPS
NIM_{t-1}	0.835 (0.000)	0.423 (0.000)	0.296 (0.021)	.292 (0.020)	.728 (0.000)	.759 (0.000)
m1			-2.24 (0.025)		-3.34 (0.001)	
m2			-0.675 (0.500)		0.4228 (0.672)	
Sargan				.178		.123

Sample: 124 Countries (1991-2000).

-Year Dummies included in all models.
-m1 and m2 are test for first and second order serial correlation for first-differenced residuals, asymptotically N (0, 1).
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2.
-P- values are reported in parentheses.

Table 2: Basic Model

Dependent Variable: Net Interest Margins (NIM_t)

	GMM DIF 1STEP	GMM DIF 2STEPS	GMM SYS 1STEP	GMM SYS 2STEPS
NIM_{t-1}	-0.068 (0.680)	0.013 (0.927)	.678 (0.000)	.720 (0.000)
$GROWTH_t$	-0.212 (0.004)	-0.159 (0.005)	-0.108 (0.009)	-0.095 (0.014)
.m1	-2.795 (0.005)		-3.19 (0.001)	
.m2	-1.591 (0.112)		0.2198 (0.826)	
Sargan		0.117		0.103

Sample: 115 Countries (1991-2000).
-Year Dummies included in all models.
-m1 and m2 are test for first and second order serial correlation, asymptotically N (0,1). These test the first-differenced residuals.
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2.
-P- values are reported in parentheses.

Table 3: Basic Model (Controlling for Financial Development).

Dependent Variable: Net Interest Margins (NIM_t)

	GMM DIF 1STEP	GMM DIF 2STEPS	GMM SYS 1STEP	GMM SYS 2STEPS
NIM_{t-1}	0.015 (0.904)	0.040 (0.804)	0.574 (0.000)	.573 (0.000)
$GROWTH_t$	-0.127 (0.007)	-0.130 (0.029)	-0.074 (0.090)	-0.077 (0.088)
$PRIV.CRED.(avg)_t$	-0.042 (0.122)	-0.033 (0.316)	-0.031 (0.000)	-0.030 (0.000)
.m1	-2.543 (0.011)		-3.04 (0.002)	
.m2	-1.553 (0.121)		0.515 (0.607)	
Sargan		0.141		0.271

Sample: 109 Countries (1991-2000).
-Year Dummies included in all models.
-m1 and m2 are test for first and second order serial correlation, asymptotically N (0,1). These test the first-differenced residuals.
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2.
-P- values are reported in parentheses.

Table 4: The Role of Concentration.

Dependent Variable: Net Interest Margins (NIM_t)

	GMM DIF 1STEP	GMM DIF 2STEPS	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS ⊥DEV
NIM_{t-1}	0.048 (0.643)	0.071 (0.598)	0.540 (0.000)	0.543 (0.000)	0.556 (0.000)
$GROWTH_t$	-0.096 (0.017)	-0.095 (0.055)	-0.083 (0.027)	-0.081 (0.030)	-0.071 (0.036)
$PRIV.CRED.(avg)_t$	-0.037 (0.228)	-0.033 (0.325)	-0.034 (0.000)	-0.033 (0.000)	-0.032 (0.000)
$CONCENTRATION_t$	-0.014 (0.571)	-0.015 (0.547)	-0.019 (0.066)	-0.017 (0.084)	-0.018 (0.062)
.m1	-2.600 (0.009)		-3.117 (0.002)		
.m2	-1.387 (0.166)		0.496 (0.620)		
Sargan		0.251		0.885	0.860

Sample: 109 Countries (1991-2000)
Year Dummies included in all models.
.m1 and .m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals.
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 5: The Role of Inflation, Real Rates and Operational Costs.

Dependent Variable: Net Interest Margins (NIM_t)

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS ⊥DEV	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS ⊥DEV
NIM_{t-1}	0.388 (0.000)	0.394 (0.000)	0.393 (0.000)	0.378 (0.000)	0.394 (0.000)	0.462 (0.000)
$GROWTH_t$	-0.078 (0.042)	-0.077 (0.051)	-0.092 (0.012)	-0.094 (0.020)	-0.092 (0.034)	-0.075 (0.046)
$PRIV.CRED.(avg)_t$	-0.022 (0.001)	-0.021 (0.001)	-0.023 (0.000)	-0.050 (0.001)	-0.048 (0.002)	-0.040 (0.000)
$CONCENTRATION_t$	-0.009 (0.286)	-0.010 (0.265)	-0.010 (0.220)	-0.021 (0.136)	-0.024 (0.093)	-0.013 (0.215)
$OVERCOSTS_t$	0.460 (0.000)	0.454 (0.000)	0.438 (0.000)	---	---	---
$INFLATION_t$	---	---	---	-0.015 (0.257)	-0.016 (0.205)	-0.012 (0.310)
$REALRATE_t$	---	---	---	0.0003 (0.312)	0.0003 (0.312)	0.0002 (0.454)
.m1	-2.934 (0.003)			-2.452 (0.014)		
.m2	0.6728 (0.501)			1.137 (0.256)		
Sargan		0.725	0.784		0.975	0.977

Sample: 109 Countries. 1991-2000
-Year Dummies included in all models.
-m1 and m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals.
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 6-A: Entry and the counter-cyclicality of the margins.

Dependent Variable: Net Interest Margins (NIM_t)

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS ⊥ DEV	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS ⊥ DEV
NIM_{t-1}	0.496 (0.001)	0.479 (0.001)	0.499 (0.000)	0.510 (0.000)	0.503 (0.000)	0.505 (0.000)
$GROWTH_t$	-0.026 (0.254)	-0.029 (0.186)	-0.024 (0.389)	-0.026 (0.343)	-0.029 (0.302)	-0.032 (0.213)
$PRIV.CRED.(avg)_t$	-0.033 (0.000)	-0.033 (0.000)	-0.032 (0.000)	-0.028 (0.001)	-0.029 (0.001)	-0.028 (0.001)
$CONCENTR._t$	-0.006 (0.636)	-0.007 (0.621)	-0.003 (0.248)	-0.003 (0.833)	-0.003 (0.799)	-0.004 (0.763)
$INFLATION_t$	0.021 (0.302)	0.022 (0.275)	0.016 (0.436)	-0.022 (0.233)	-0.024 (0.231)	-0.026 (0.268)
$REALRATE_t$	0.0001 (0.359)	0.0002 (0.360)	0.0002 (0.312)	0.0002 (0.260)	0.0002 (0.238)	0.0002 (0.248)
$Foreign\ Banks_{t-1}$	-0.034 (0.175)	-0.033 (0.188)	-0.031 (0.214)	----	----	
$Foreign\ Banks_{t-2}$	0.026 (0.027)	0.025 (0.032)	0.026 (0.037)	----	----	
$\Delta Foreign\ Banks_{t-1}$	----	----	----	-0.031 (0.046)	-0.032 (0.068)	-0.033 (0.063)
.m1	-2.107 (0.035)			-2.452 (0.014)		
.m2	1.537 (0.124)			1.137 (0.256)		
Sargan		1.000	0.784		1.000	1.000

Sample: 95 Countries (1991-2000) -Year Dummies included in all models. - m1 and m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals. -The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 6-B: Entry and the counter-cyclicality of the margins.

Dependent Variable: Net Interest Margins (NIM_t)

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV
NIM_{t-1}	0.341 (0.019)	0.348 (0.001)	0.389 (0.005)	0.342 (0.020)	0.351 (0.000)	0.390 (0.005)
$GROWTH_t$	0.021 (0.703)	0.023 (0.695)	-0.021 (0.683)	0.013 (0.825)	-0.010 (0.869)	-0.030 (0.586)
$PRIV.CRED.(avg)_t$	-0.014 (0.068)	-0.012 (0.150)	-0.015 (0.265)	-0.013 (0.069)	-0.011 (0.115)	-0.014 (0.033)
$CONCENTR._t$	-0.019 (0.174)	-0.016 (0.251)	-0.014 (0.265)	-0.015 (0.311)	-0.012 (0.398)	-0.012 (0.383)
$OVERCOSTS_t$	0.500 (0.010)	0.520 (0.009)	0.458 (0.009)	0.516 (0.005)	0.520 (0.006)	0.467 (0.006)
$Foreign\ Banks_{t-1}$	-0.042 (0.126)	-0.036 (0.067)	-0.040 (0.159)	----	----	----
$Foreign\ Banks_{t-2}$	0.030 (0.098)	0.027 (0.099)	0.031 (0.084)	----	----	----
$\Delta Foreign\ Banks_{t-1}$	----	----	----	-0.033 (0.093)	-0.031 (0.115)	-0.033 (0.094)
.m1	-2.792 (0.005)			-2.801 (0.005)		
.m2	2.017 (0.044)			2.028 (0.043)		
Sargan		0.464	0.507		0.375	1.000

Sample: 95 Countries (1991-2000). -Year Dummies included in all models. -m1 and .m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals. -The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 7-A - Regional Analysis- Basic Model-

Dependent Variable: Net Interest Margins (NIM_t)

	GMM DIF 1STEP	GMM DIF 2STEPS	GMM SYS 1STEP	GMM SYS 2STEPS
$NIM_{t-1} * POOR$	0.070 (0.601)	0.098 (0.515)	.664 (0.000)	.678 (0.000)
$NIM_{t-1} * (1 - POOR)$	1.15 (0.001)	1.06 (0.002)	0.403 (0.001)	0.441 (0.002)
$GROWTH_t * POOR$	-0.207 (0.002)	-0.163 (0.017)	-0.113 (0.013)	-0.111 (0.030)
$GROWTH_t * (1 - POOR)$	-0.057 (0.316)	-0.033 (0.542)	-0.078 (0.081)	-0.076 (0.105)
.m1	-3.519 (0.005)		-3.18 (0.001)	
.m2	-1.462 (0.144)		0.2084 (0.835)	
Sargan		0.697		0.773

Sample: 115 Countries (1991-2000).
-Year Dummies included in all models.
-m1 and .m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals.
-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 7-B -Regional Analysis- Extended Model-

Dependent Variable: Net Interest Margins (NIM_t)

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV
$NIM_{t-1} * POOR$	0.566 (0.000)	0.576 (0.000)	0.590 (0.000)	0.584 (0.000)	0.630 (0.000)	0.588 (0.000)
$NIM_{t-1} * (1-POOR)$	0.504 (0.000)	0.547 (0.000)	0.540 (0.000)	0.562 (0.000)	0.613 (0.000)	0.568 (0.000)
$GROWTH_t * POOR$	-0.064 (0.082)	-0.066 (0.101)	-0.052 (0.109)	-0.075 (0.056)	-0.074 (0.103)	-0.080 (0.037)
$GROWTH_t * (1-POOR)$	-0.055 (0.115)	-0.015 (0.708)	-0.059 (0.099)	-0.067 (0.092)	-0.019 (0.674)	-0.072 (0.383)
$PRIV.CRED(avg) * POOR$	-0.032 (0.067)	-0.029 (0.081)	-0.028 (0.051)	-0.033 (0.021)	-0.022 (0.036)	-0.030 (0.006)
$PRIV.CRED(avg) * (1-POOR)$	-0.025 (0.004)	-0.027 (0.011)	-0.023 (0.003)	-0.025 (0.009)	-0.025 (0.007)	-0.025 (0.006)
$CONCENTR. * POOR$	----	----	----	-0.0043 (0.608)	-0.053 (0.469)	-0.0058 (0.456)
$CONCENTR. * (1-POOR)$	----	----	----	-0.011 (0.259)	-0.063 (0.541)	-0.012 (0.234)
.m1	-3.041 (0.002)			-3.195 (0.001)		
.m2	0.5090 (0.611)			0.5798 (0.562)		
Sargan		1.000	1.000		1.000	1.000

Sample: 109 Countries (1991-2000). -Year Dummies included in all models. - .m1 and .m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals.-The Sargan Test for overidentifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 8: Entry and the counter-cyclicality of the margins (Regional Analysis)

Dependent Variable: Net Interest Margins (NIM_t)

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV
NIM_{t-1}	0.253 (0.087)	0.274 (0.078)	0.385 (0.005)
$GROWTH_t$	-0.012 (0.673)	-0.028 (0.368)	-0.006 (0.800)
$PRIV.CRED.(avg)_t$	-0.039 (0.001)	-0.037 (0.002)	-0.033 (0.000)
$CONCENTR._t$	-0.0009 (0.951)	0.024 (0.251)	-0.001 (0.913)
$\Delta Foreign\ Banks_{t-1} * POOR$	-0.044 (0.023)	-0.041 (0.067)	-0.033 (0.073)
$\Delta Foreign\ Banks_{t-1} * (1 - POOR)$	0.020 (0.102)	0.018 (0.099)	0.032 (0.006)
.m1	-2.369 (0.018)		
.m2	1.667 (0.096)		
Sargan		0.992	0.984

Sample: 95 Countries. 1991-2000. -Year Dummies included in all models. - .m1 and .m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals.-The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 9: Determinants of entry in the local financial system

Dependent Variable: $\Delta Foreign\ Banks_t$

	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV	GMM SYS 1STEP	GMM SYS 2STEPS	GMM SYS $\perp$ DEV
$GROWTH_t$	0.870 (0.026)	1.05 (0.036)	0.837 (0.014)	0.599 (0.097)	0.528 (0.088)	0.595 (0.081)
$GROWTH_{t-1}$	0.027 (0.057)	0.030 (0.137)	0.014 (0.520)	0.017 (0.201)	0.018 (0.134)	0.012 (0.369)
$RULE\ of\ LAW$	-0.001 (0.818)	-0.0005 (0.933)	-0.0004 (0.954)	0.013 (0.136)	0.011 (0.150)	0.012 (0.190)
$N_o.\ of\ REVOLUTIONS$	-0.012 (0.635)	-0.007 (0.798)	-0.011 (0.605)	-0.054 (0.104)	-0.048 (0.876)	-0.051 (0.131)
$FREEDOM$	-0.012 (0.429)	-0.013 (0.480)	-0.008 (0.678)	-0.023 (0.232)	-0.022 (0.179)	-0.024 (0.234)
$PRIV.CRED.(avg)_t$	----	----	----	-0.014 (0.734)	-0.059 (0.876)	-0.002 (0.995)
$CONCENTR._t$	----	----	----	-0.012 (0.084)	-0.010 (0.150)	-0.011 (0.137)
GDP_t	----	----	----	0.044 (0.038)	0.039 (0.061)	0.040 (0.069)
.m1	-2.369 (0.018)			-3.096 (0.002)		
.m2	1.667 (0.096)			0.8451 (0.398)		
Sargan		0.817	0.912		1.000	1.000

Sample: 67 Developing Countries (1991-2000) -Year Dummies included in all models. - m1 and m2 are test for first and second order serial correlation, asymptotically N (0, 1). These test the first-differenced residuals. -The Sargan Test for over-identifying restrictions for the GMM estimators is asymptotically χ^2. (P- values are reported in parentheses.)

Table 10: Standard Deviation

VARIABLE	RBC	ICFM
Output	1.80	2.31
Consumption	0.52	0.70
Investment	5.74	12.08
Capital	0.49	0.94
Employment	1.37	2.19

Notes: Theoretical second moments (as percentage deviation from steady state values) are reported. RBC refers to the standard RBC model and ICFM to the monopolistic financial market setup respectively. The method used was the frequency domain technique depicted in Uhlig (1999). The series are H-P filtered with a smoothness parameter of 1600 so that only cyclical components remain.

Table 11: Relative Standard Deviation

VARIABLE	RBC	ICFM
Output	1.00	1.00
Consumption	0.29	0.30
Investment	3.19	5.23
Capital	0.28	0.41
Employment	0.76	0.95

Notes: Relative standard deviations respect to output. For further information, see notes to Table 10.

Table 12: Sensitivity Analysis

VARIABLE	$\tau = 0.20$	$\tau = 0.40$	$\tau = 0.60$	$\tau = 0.80$
Output	1.84	2.02	2.21	2.41
Consumption	0.60	0.62	0.66	0.74
Investment	8.24	9.67	11.26	12.89
Capital	0.68	0.78	0.89	0.98
Employment	1.43	1.71	2.03	2.35

Notes: Implied volatilities for different parameter values of τ . For further information, see notes to Table 10.

Appendix 2
Variable Definitions

NIM: Net interest income minus interest over total assets.

GROWTH: Annual growth rate of real GDP.

PRIV. CREDIT.: Private Credit by deposit money banks to GDP, calculated using the following deflation method: $\{(0.5)*[F_t / P_{et} + F_{t-1} / P_{et-1}]\}/[GDP_t / a_t]$

CONCENTRATION: Assets of three largest banks as a share of assets of all commercial banks in the system.

OVERCOSTS: Accounting value of a bank's overhead costs as a share of its total assets.

INFLATION: Annual inflation from the GDP deflator.

REALRATE: Real Interest Rate.

Foreign Banks: Number of foreign banks to total number of banks. A bank is defined to be a foreign bank if it has at least fifty percent of foreign ownership.

Δ *Foreign Banks*: Variation in the number of foreign banks.

GDP: Real GDP per capita.

No. of Revolutions: Average number of revolutions (1970-2000).

FREEDOM: Freedom House Ratings.

Rule of Law: ICRG Law and Order Rating.