An Internally Consistent Approach to the Estimation of Market Power and Cost Efficiency with an Application to U.S. Banking*

Efthymios G. Tsionas[1] Emir Malikov[2] Subal C. Kumbhakar[3,4]

[1]Department of Economics, Lancaster University Management School, Lancaster, United Kingdom
[2]Department of Agricultural Economics, Auburn University, Auburn, AL, United States
[3]Department of Economics, State University of New York at Binghamton, Binghamton, NY, United States
[4]University of Stavanger Business School, Stavanger, Norway

January 25, 2018

Abstract

We develop a novel unified econometric methodology for the formal examination of the market power – cost efficiency nexus. Our approach can meaningfully accommodate a mutually dependent relationship between the firm's cost efficiency and marker power (as measured by the Lerner index) by explicitly modeling the simultaneous determination of the two in a system of nonlinear equations consisting of the firm's cost frontier and the revenue-to-cost ratio equation derived from its stochastic revenue function. Our framework places no a priori restrictions on the sign of the dependence between the firm's market power and efficiency as well as allows for different hierarchical orderings between the two, enabling us to discriminate between competing quiet life and efficient structure hypotheses. Among other benefits, our approach completely obviates the need for second-stage regressions of the cost efficiency estimates on the constructed market power measures which, while widely prevalent in the literature, suffer from multiple econometric problems as well as lack internal consistency/validity. We showcase our methodology by applying it to a panel of U.S. commercial banks in 1984–2007 using Bayesian MCMC methods.

Keywords: Productivity & Competitiveness, Efficiency, Market Power, Lerner Index, Banks, Quiet Life Hypothesis

JEL Classification: C11, C30, D24, D40, G21

Email: m.tsionas@lancaster.ac.uk (Tsionas), emalikov@auburn.edu (Malikov), kkar@binghamton.edu (Kumbhakar).

We would like to thank the editor, Robert Dyson, and three anonymous referees for many insightful comments that helped improve this paper. Any remaining errors are our own.

1 Introduction

Owing to its important social welfare implications, researchers have long been interested in disentangling the complex relationship between the market structure and the cost efficiency of firms. The two notable (structural) hypotheses put forth to conceptualize this relationship — the quiet life and the efficient structure hypotheses (thereafter, QLH and ESH) — contrast rather sharply both in the sign and implied directionality of the relationship. Not only is it still unclear if the relationship is positive or negative, but it also remains unsettled whether the market structure determines the firms' performance, including their efficiency, or whether the market structure should rather be viewed as an endogenous outcome of firms' behavior reflective (at least partly) of their efficiency levels.

The QLH as postulated by Hicks (1935) conceptualizes the market structure as a determinant of firms' efficiency, whereby firms with higher market power trade potential monopoly rents for lower efficiency. This negative relationship may exist because, owing to the high levels of market power providing them with the "price cushion", firm managers might not work as hard to keep costs at minimum or might expend resources to obtain/maintain the market power, i.e., engage in further rent-seeking (Berger & Hannan, 1998). In contrast, Demsetz's (1973) ESH asserts that the industry market structure is instead an outcome of the interaction among individual firms exhibiting different efficiency levels, through which more efficient firms gain larger market shares and hence secure greater monopolistic power. Such a positive relationship between the cost efficiency and market power is oftentimes rationalized as the result of more efficient firms with superior management out-competing their less efficient rivals which operate at higher costs (Berger, 1995).

Empirical work on the nexus between firm efficiency and market power has particularly favored the U.S. banking sector as a "laboratory" for analysis owing to the relative homogeneity of banks in the industry which helps facilitate cross-firm performance comparisons (e.g., Berger & Hannan, 1998; Jayaratne & Strahan, 1998; Koetter, Kolari & Spierdijk, 2012). The findings, however, have been rather mixed. While Berger & Hannan (1998) and Delis & Tsionas (2009) find that, consistent with the QLH, U.S. banks exhibiting greater market power tend to suffer from significant cost efficiency losses, the finding of a positive relationship between cost efficiency and market power of European banks documented by Weill (2004), Casu & Girardone (2006) and Maudos & Fernández de Guevara (2007) instead buttress the ESH. More recently, Koetter et al. (2012) have also reported the empirical evidence pointing to a positive effect of the market power on the bank's cost efficiency in the U.S. which lends support to ESH.[1] To the contrary, in line with QLH, Koetter & Vins (2008) consistently find a negative relationship between bank-level measures of market power and cost efficiency for German savings banks. Similar findings are reported by Turk Ariss (2010) and Dong, Firth, Hou & Yang (2016) for banks in developing countries including China.

While different in their choice of the measure of market power and in some other methodological aspects, overwhelming majority of such studies resort to a two-stage analysis which suffers from multiple fundamental econometric problems (discussed below) casting a serious shadow on the validity and reliability of their findings. The same concerns also apply to a broader literature examining the links between the market structure/competition and firms' profitability, profit efficiency or stability,[2] which the papers on the cost efficiency – market power nexus closely relate

[1] They do document an opposite finding in support of the QLH using measures of *profit* efficiency. Restrepo-Tobón & Kumbhakar (2014) however cast doubt on the latter finding in their re-examination of Koetter et al. (2012).

[2] Examples of such studies in banking (with oftentimes contrary findings) include Molyneux, Lloyd-Williams & Thornton (1994), Goldberg & Rai (1996), Punt & van Rooij (1999), Claessens & Laeven (2005), de Guevara & Maudos (2007), Schaeck & Cihák (2008), Koetter & Poghosyan (2009), Carbo, Humphrey, Maudos & Molyneux (2009), among many others.

to and share much of their methodology with. In this paper, we contribute to the literature by proposing a novel unified econometric approach that tackles the problems associated with such two-stage analyses widely favored in the literature.

To make matters more concrete, we focus on the estimation issues concerning the examination of the *firm-level* relationship between the market power and cost efficiency. Thus, we measure firms' market power using the Lerner index which is a popular go-to firm-specific measure of monopolistic power in the literature (e.g., de Guevara & Maudos, 2007; Berger, Klapper & Turk-Ariss, 2009; Koetter et al., 2012; Das & Kumbhakar, 2016) since the alternative indices such as Herfindahl-type concentration ratios or the Rosse-Panzar H-statistic usually yield industry-level estimates[3] of competitiveness in the market (for more on these measures, e.g., see Bolt & Humphrey, 2015). The cost efficiency estimates are obtained using a widely popular stochastic frontier formulation of the cost function (e.g., Berger & Mester, 1997, 2003; Malikov, Kumbhakar & Tsionas, 2016). Traditionally, researchers perform their analyses in two stages, where they first compute the Lerner index using the fitted scale elasticity of cost as well as estimate firm-specific cost efficiency scores and then regress the obtained cost efficiency estimates on the Lerner index in the second stage (e.g., see Maudos & Fernández de Guevara, 2007; Koetter & Vins, 2008; Turk Ariss, 2010; Koetter et al., 2012). However, not only does the latter two-stage model lack internal consistency (validity) but it also suffers from a number of acute econometric issues.

Specifically, the two-stage analysis fails to accommodate a simultaneous determination of the Lerner index (firm's market power) and the cost efficiency. As discussed earlier, the directionality of the relationship between the market power and efficiency is ambiguous, and causality likely runs both ways, whereby more efficient firms are able to survive the competition and thus acquire greater market power which in turn may create "quiet life" incentives leading to a decline in firms' efficiency. Even if this "reverse causality" problem is acknowledged by researchers, the methodology used to tackle it however usually falls short of its overreaching task. The endogeneity is oftentimes argued to be resolved either by merely "adjusting" the Lerner index whereby the efficiency-corrected estimates of the marginal cost are used in the computation of the index (e.g., Turk Ariss, 2010) and/or by employing instruments in the second-stage regressions of the efficiency estimates on the market power index (e.g., Berger & Hannan, 1998; Berger et al., 2009; Koetter et al., 2012). However, neither of these methods can meaningfully accommodate the simultaneity of the firm's market power and efficiency. While the "adjusted" Lerner index indeed explicitly acknowledges the existence of cost inefficiency, in its construction however, researchers use the marginal cost estimates from the stochastic cost frontier model that assumes complete independence of the cost (in)efficiency from the cost function covariates and hence the Lerner index. That is, the efficiency-adjusted Lerner index is "adjusted" under the assumption that cost efficiency is *independent* from the firm's market power. Not only may the cost (in)efficiency be severely biased because one forcefully imposes its independence from the market power during the estimation (Delis & Tsionas, 2009), but this independence also suggests that any second-stage regressions of the cost efficiency on the market power indicators (and possibly other contextual variables) contradict the underlying assumption of the first-stage regression and thus are likely to be spurious. Worse yet, the cost inefficiency is oftentimes assumed to be not only independently but also *identically* distributed across firms with constant mean and variance (e.g., Maudos & Fernández de Guevara, 2007; Turk Ariss, 2010; Koetter et al., 2012), which implies no systemic relationship between the firm's efficiency and other covariates thus rendering any second-stage analysis void. Subsequently, the two-stage methodology lacks internal consistency (validity) by which both stages would be reconcilable with one another.

[3]Brissimis & Delis (2011) have recently proposed employing nonparametric local regression techniques to obtain estimates of the Rosse-Panzar H-statistic at a finer level such as the level of a unit.

In its second stage, virtually no study recognizes that both the firm's efficiency and the Lerner index are in fact generated estimates subject to parameter uncertainty which needs to be accounted for when performing inference. While the sampling uncertainty associated with the first-stage estimate used as a left-hand-side variable in the second-stage analysis (usually, cost efficiency) does not generally pose a problem, the same cannot be said about the generated regressor (usually, the Lerner index) used as a right-hand-side explanatory variable. Further, the standard approach, whereby the Lerner index is constructed using raw information on the firm's revenues and estimates of the marginal cost from a stochastic cost frontier, implicitly assumes away any stochasticity in the firm's revenue function. The latter is rather arbitrary; it only appears logical to allow for stochastic noise in both the firm's costs *and* revenues. Lastly but not least importantly, the second-stage regressions of the cost efficiency scores on the Lerner index are usually estimated via least squares (ordinary or two-stage) without taking a bounded codomain of the cost efficiency estimates (or the log thereof) into consideration (e.g., Berger & Hannan, 1998; Weill, 2004; Koetter et al., 2012).[4]

In this paper, we seek to offer a solution to the above econometric problems associated with the formal examination of the cost efficiency – market power nexus. More specifically, we propose a novel, internally consistent approach to modeling a mutually dependent relationship between the firm's cost efficiency and marker power (as measured by the Lerner index), which explicitly accommodates the simultaneous/endogenous determination of the two and completely obviates the need for a second-stage analysis. Both the firm's cost efficiency and marker power index are estimated jointly and derived from a single unified model thus enabling us to interpret and analyze them on a common ground.

We begin by recognizing that neither the cost efficiency nor the market power are observed and, therefore, ought to be treated accordingly when estimating the firm's production process. We let both of these latent variables directly co-depend in a variety of ways, thereby having efficiency be automatically adjusted for market power and vice versa, in a system of nonlinear equations consisting of the firm's cost frontier and the revenue-to-cost ratio equation derived from the firm's stochastic revenue function. Our framework places no *a priori* restrictions on the sign of the dependence between market power and efficiency as well as allows for different hierarchical orderings between the two, enabling us to meaningfully discriminate between the QLH and ESH. To draw statistical inference, we consider three alternative econometric specifications of our unified system-based model which we estimate using Markov Chain Monte Carlo (MCMC) methods.

We showcase our unified model by applying it to a panel of U.S. commercial banks operating in 1984–2007. Regardless of the econometric specification of the model used, the data consistently point to a negative dependence between the bank's cost efficiency and the Lerner index thus providing empirical evidence in support of the QLH. This finding is reversed when we employ a traditional two-stage analysis where both the cost efficiency and the adjusted Lerner index are estimated separately without allowing for a simultaneous determination of the two. The latter highlights the pivotal importance of a proper econometric modeling of the market power – efficiency relationship which a popular two-stage analysis is unable to deliver.

The rest of the paper proceeds as follows. Section 2 introduces our unified model of market power and efficiency. We describe three alternative econometric specifications of our model in Section 3 with the details relegated to the Appendix. Section 4 reports the empirical application. We conclude in Section 5.

[4]Few exceptions include Maudos & Fernández de Guevara (2007) and Turk Ariss (2010) who estimate logistic and tobit regressions in their second-stage analyses, respectively.

2 A Unified Model

In this section, we propose a novel, internally consistent approach to modeling a mutually dependent relationship between the firm's cost efficiency and marker power (as measured by the Lerner index), which explicitly accommodates the simultaneous/endogenous determination of the two.

For the ease of exposition, we first consider the case where a firm, which may potentially exercise monopoly power in the market, produces a *single* output. Suppose the firm's cost function is given by $C = C(\mathbf{w}, y) \equiv \min_{\mathbf{x}} \{\mathbf{w}'\mathbf{x} \mid y \leq F(\mathbf{x})\} : \mathbb{R}^J_{++} \times \mathbb{R}_+ \to \mathbb{R}_+$, where $\mathbf{x} \in \mathbb{R}^J_+$ is the vector of inputs with the corresponding vector of input prices $\mathbf{w} \in \mathbb{R}^J_{++}$, y is the total output quantity, and $F(\mathbf{x})$ is the production function. Further, let the output price be denoted by $p \in \mathbb{R}_{++}$. To measure the firm's market power, we use the Lerner index widely favored in the literature. Specifically, we define the Lerner index of market power as follows: $\widetilde{L} = \big(p - \frac{\partial C(\mathbf{w},y)}{\partial y}\big)/p \in [0, 1)$, where $\frac{\partial C(\mathbf{w},y)}{\partial y}$ is the firm's marginal cost. The rationale of the index is that the monopolistic firm can set the price above the marginal cost (which normally equals the competitive price) and, therefore, the excess of price over marginal cost should be a good measure of market power. Higher values of the index imply greater market power, whereas the zero value (a lower boundary) points to a perfectly competitive firm. Lastly, note that the Lerner index is closely related to the conventional measure of the firm's markup which uses the marginal cost, as opposed to the output price, as the reference for sizing the firm's monopoly power, namely $\big(p - \frac{\partial C(\mathbf{w},y)}{\partial y}\big)\big/\frac{\partial C(\mathbf{w},y)}{\partial y}$.

We next rewrite the Lerner index $\widetilde{L}$ as a function of the firm's revenue and the output elasticity of its cost. Specifically,

$$\widetilde{L} = \frac{p - \frac{\partial C(\mathbf{w},y)}{\partial y}}{p} = \frac{p - \frac{C}{y} \times \varepsilon_y}{p} = \frac{R - C \times \varepsilon_y}{R}, \tag{2.1}$$

where $\varepsilon_y = \frac{\partial \ln C(\mathbf{w},y)}{\partial \ln y}$ is the output (scale) elasticity of cost, and $R = py \in \mathbb{R}_+$ is the total revenue.

From (2.1), it is immediately evident that, had the output elasticity of cost ε_y been known and/or directly observable from the data, we could have easily obtained the estimate of the Lerner index. However, given the unobservability of ε_y, we need to estimate it too. To do so, we can make use of the information from the firm's cost function.[5]

Before proceeding to the details of how the output elasticity of cost is estimated, we first generalize the above discussion to the case when the firm engages in a *multi*-output production. The firm's multi-output cost function can then be redefined as

$$C = C(\mathbf{w}, \mathbf{y}) \equiv \min_{\mathbf{x}} \big\{\mathbf{w}'\mathbf{x} \mid T(\mathbf{x}, \mathbf{y}) \geq 1\big\} : \mathbb{R}^J_{++} \times \mathbb{R}^M_+ \to \mathbb{R}_+, \tag{2.2}$$

where $\mathbf{y} \in \mathbb{R}^M_+$ is the vector of outputs, $T(\cdot)$ is the transformation function relating inputs to outputs within the set of technologically feasible combinations $\mathbb{T} = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \text{ can produce } \mathbf{y}\}$, and $\mathbf{x}$ and $\mathbf{w}$ are just as defined earlier. The vector of output prices is now given by $\mathbf{p} \in \mathbb{R}^M_{++}$.

In the instance of M outputs being produced by the firm, it is now imperative to recognize that, following our earlier definition, one can define M different output-specific Lerner indices, i.e., $\widetilde{L}_m = \big(p_m - \frac{\partial C(\mathbf{w},\mathbf{y})}{\partial y_m}\big)/p_m \;\forall\; m = 1, \ldots, M$, with the corresponding M output-specific markup measures. To get around this issue, we therefore employ a multi-output definition of the Lerner index where the market power is gauged on the basis of the differential between the firm's average total revenue (as opposed to the price) and its "average" marginal cost (e.g., Koetter & Poghosyan,

[5]See Delis, Iosifidi & Tsionas (2014), for an excellent discussion of the estimation of marginal cost.

2009; Koetter et al., 2012; Das & Kumbhakar, 2016). Similar to (2.1), the multi-output Lerner index can then be written as

$$L = \frac{R - C \times \sum_m \varepsilon_{y_m}}{R}, \tag{2.3}$$

where $\varepsilon_{y_m} = \frac{\partial \ln C(\mathbf{w},\mathbf{y})}{\partial \ln y_m}$ is the elasticity of cost with respect to the mth output, and $R = \sum_m p_m y_m \in \mathbb{R}_+$ is the total revenue as defined earlier. Essentially, such an index provides a measure of the firm's "average" monopolistic power across the output markets. From (2.3), it follows that

$$\frac{R}{C} = \frac{1}{1-L} \times \sum_m \varepsilon_{y_m}, \tag{2.4}$$

which in logs yields

$$\ln\left(\frac{R}{C}\right) = \ln\left(\sum_m \varepsilon_{y_m}\right) + u_L, \tag{2.5}$$

where, given the permissible range of the Lerner index, we define the (latent) unbounded one-sided transformation of L as $u_L = -\ln(1-L) \geq 0$ which is increasing in the firm's marker power.

Since the so-called "scale elasticity" $\sum_m \varepsilon_{y_m}$ is unobservable for a direct computation of u_L (and hence of L), we recover it from the firm's dual cost function which we estimate simultaneously along with equation (2.5) while also allowing for *(i)* the presence of one-sided cost inefficiency and *(ii)* mutual dependence between the latter and the Lerner index. Formally, after appending both equations with two-sided stochastic errors, the nonlinear system of simultaneous equations, which we seek to estimate, is given by

$$\ln C = \ln \mathcal{C}(\mathbf{w}, \mathbf{y}; \boldsymbol{\beta}) + u_C + v_C \tag{2.6a}$$

$$\ln\left(\frac{R}{C}\right) = \ln\left(\sum_m \varepsilon_{y_m}(\mathbf{w}, \mathbf{y}; \boldsymbol{\beta})\right) + u_L + v_L, \tag{2.6b}$$

where $\ln \mathcal{C}(\cdot\,; \boldsymbol{\beta})$ is some parametric specification of the unknown cost function with $\boldsymbol{\beta}$ being the vector of unknown parameters, and $u_C \geq 0$ is the one-sided cost *in*efficiency term. Here, we explicitly recognize that ε_{y_m} is a function of the unknown parameters $\boldsymbol{\beta}$, since the output elasticity of cost is found as the partial log-derivative of the fitted cost function, i.e., $\varepsilon_{y_m} = \frac{\partial \ln \mathcal{C}(\mathbf{w},\mathbf{y};\boldsymbol{\beta})}{\partial \ln y_m}$ $\forall\, m = 1, \ldots, M$. Following the popular practice, we use the translog specification for $\ln \mathcal{C}(\cdot\,; \boldsymbol{\beta})$, known to yield a flexible second-order approximation to an arbitrary, unknown functional form for the cost function.

Our system-of-equations approach presents a superior alternative to the standard practice of empirical assessment of the efficiency – market power relationship in the literature, whereby one first estimates the firm's cost function to obtain estimates of cost efficiency and scale elasticity and then, as a second step, uses the latter to compute the Lerner index. Having done the above, a common researcher then proceeds to the second-stage regressions of the efficiency estimates on the Lerner index estimates and, potentially, some other covariates. In principle, there are multiple problematic issues with this practice (as discussed in detail in Introduction) including the lack of internal consistency/validity. Therefore, to ensure the coherence and internal consistency of the econometric model used for the analysis of the market power – cost efficiency relationship, it is pivotal that one estimates the firm's level of efficiency and market power *jointly* while allowing for potential dependence not only between stochastic noises v_C and v_L, owing to the seemingly-unrelated-regressions structure of model in (2.6), but also between u_C and u_L themselves. Not only does such a joint estimation automatically "adjust" both measures, but it also completely obviates

the need for a second-stage analysis plundered by numerous econometric problems. In fact, explicit modeling of the potential dependence between u_C and u_L is the key to a consistent[6] testing of the QLH versus ESH. Unfortunately, statistical inference in this context is quite complicated due to the fact that both the cost inefficiency and the market power are unobserved latent variables which are one-sided and correlated. Thus, formal statistical discrimination between the QLH and ESH may be a rather burdensome task.

3 Econometric Specification

Our model of endogenous (simultaneous) interplay between the firm's cost efficiency and market power in (2.6) can be formalized econometrically in the number of ways. In what follow, we propose three alternative methods of modeling the market power – cost efficiency relationship, which differ in their formulation of the cross-equation dependence between (2.6a) and (2.6b). Regardless of the specifics, all three models accommodate mutual dependence of the cost inefficiency u_C and the log-Lerner-index function u_L via a simultaneous-equations structure.

In Model I, we take the most "agnostic" view of directionality of the relationship between the firm's cost efficiency and market power by imposing no structure onto the order of causality in their interdependence. We do so by modeling their joint one-sided distribution with the dependence being controlled by the covariance parameter. As an alternative, Model II postulates an *a priori* hierarchical dependence between cost efficiency and market power, whereby the latter is permitted to directly affect the latent inefficiency term via its mean. Model III relaxes Model II by explicitly allowing for *bi*directional cross-equation effects between cost efficiency and market power working through their respective means. Thus, in contrast to Model I which also accommodates bidirectional effects but does so via a *single* dependence parameter, Model III allows one to directly examine if both directions are significant (owing to the model's explicit *two*-parameter formulation of the cross-equation dependence). However, neither of the three model specifications impose restrictions onto *signs* of the parameters controlling the dependence between u_C and u_L. Ultimately, the data are to tell which model describes them more adequately.

To aid the discussion, we first rewrite our system of equations in (2.6) in a stylized form:

$$y_{1,it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{1,it} + u_{1,it} \tag{3.1a}$$

$$y_{2,it} = f(\mathbf{x}_{it};\boldsymbol{\beta}) + v_{2,it} + u_{2,it} \qquad \forall\ i = 1,\ldots,n;\ t = 1,\ldots,T, \tag{3.1b}$$

where $y_{1,it}$ and $\mathbf{x}_{it}$ denote $\ln C_{it}$ and the associated translog expansion of the cost determinants, respectively; $y_{2,it}$ denotes $\ln(R_{it}/C_{it})$; $f(\mathbf{x}_{it};\boldsymbol{\beta}) \equiv \ln\left(\sum_m \varepsilon_{y_m}\right)$ denotes the log of scale elasticity of cost defined as the sum of output elasticities derived by differentiating the translog cost function specification with respect to the log outputs; and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters to be estimated. (The details of the MCMC techniques used to estimate each of the three models are relegated to the Appendix.)

3.1 Model I: Dependence via Joint Distribution

In Model I, the cost inefficiency u_1 and the log-Lerner-index function u_2 are postulated to be mutually dependent by following a joint one-sided distribution with the dependence between the two being controlled by the covariance parameter. Formally, the stochastic assumptions for Model

[6]Both in the economic and econometric sense.

I are as follows:

$$\begin{bmatrix} v_{1,it} \\ v_{2,it} \end{bmatrix} \sim \mathbb{N}(\mathbf{0}, \mathbf{\Sigma}), \tag{3.2a}$$

$$\begin{bmatrix} u_{1,it} \\ u_{2,it} \end{bmatrix} \sim \mathbb{N}^{+}(\boldsymbol{\mu}, \mathbf{\Omega}), \ u_{1,it} \geq 0, u_{2,it} \geq 0, \tag{3.2b}$$

independently of each other as well as of the regressors $\mathbf{x}_{it}$, where $\boldsymbol{\mu} = [\mu_1, \mu_2]'$ and $\mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}$.

For the ease of exposition, in the remainder of this subsection, we treat (μ_1, μ_2) as constants. However, we can easily allow the means of $u_{1,it}$ and $u_{2,it}$ to vary with some contextual covariates via the following parameterization: $\mu_{1,it} = \mathbf{z}'_{1,it}\boldsymbol{\gamma}_1$ and $\mu_{2,it} = \mathbf{z}'_{2,it}\boldsymbol{\gamma}_2$, where $\mathbf{z}_{1,it}$ and $\mathbf{z}_{2,it}$ is an $l_1 \times 1$ and $l_2 \times 1$ vector of covariates, and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ is an $l_1 \times 1$ and $l_2 \times 1$ vector of the corresponding parameters, respectively. In fact, we do so in the empirical application, where we let μ_1 and μ_2 be time-varying and firm-specific by conditioning them on heterogeneous firm characteristics.

Next, we write $\mathbf{x}_{it} = \left[\mathbf{x}'_{o,it}, \mathbf{x}'_{*,it}\right]'$, where $\mathbf{x}_{o,it}$ is a $k_o \times 1$ vector corresponding to the linear terms in the translog expansion. The scale elasticity $\sum_m \varepsilon_{y_{m,it}}$ can then be computed as $\sum_m \varepsilon_{y_{m,it}} = \mathbf{x}'_{o,it}\mathbf{R}\boldsymbol{\beta}$, where $\mathbf{R}$ is a $k_o \times k$ selection matrix whose elements are either 0 or 1. Also, for convenience, define $\mathbf{y}_{it} = [y_{1,it}, y_{2,it}]'$, $\mathbf{v}_{it} = [v_{1,it}, v_{2,it}]'$ and $\mathbf{u}_{it} = [u_{1,it}, u_{2,it}]'$.

The conditional distribution of the observables is given by

$$p(\mathbf{y}_{it}|\mathbf{x}_{it}, \boldsymbol{\theta}) = (2\pi)^{-1} C_\Omega |\mathbf{\Sigma}|^{-1/2} \times \int_0^\infty \int_0^\infty \exp\left\{ -\frac{1}{2}(\mathbf{r}_{it} - \mathbf{u}_{it})' \mathbf{\Sigma}^{-1} (\mathbf{r}_{it} - \mathbf{u}_{it}) - \frac{1}{2}(\mathbf{u}_{it} - \boldsymbol{\mu})' \mathbf{\Omega}^{-1} (\mathbf{u}_{it} - \boldsymbol{\mu}) \right\} d\mathbf{u}_{it}, \tag{3.3}$$

where $\mathbf{r}_{it} = \mathbf{y}_{it} - \begin{bmatrix} \mathbf{x}'_{it}\boldsymbol{\beta} \\ f(\mathbf{x}_{it}; \boldsymbol{\beta}) \end{bmatrix}$, $C_\Omega = (2\pi)^{-1} |\mathbf{\Omega}|^{-1/2} \Phi_\rho \left(\frac{\mu_1}{\omega_{11}}, \frac{\mu_2}{\omega_{22}} \right)^{-1}$ is the normalizing constant of the bivariate truncated normal distribution with Φ_ρ denoting the bivariate standard normal distribution function with the correlation coefficient $\rho = \omega_{12} (\omega_{11}\omega_{22})^{-1/2}$, and $\boldsymbol{\theta} = \left[\boldsymbol{\beta}', vech(\mathbf{\Sigma})', vech(\mathbf{\Omega})', \boldsymbol{\mu}'\right]'$ denotes the collective vector of all parameters.

It is straightforward but tedious to generalize Pitt & Lee (1981, Appendix 2) and express (3.3) in the following form:

$$p(\mathbf{y}_{it}|\mathbf{x}_{it}, \boldsymbol{\theta}) \propto |\mathbf{\Omega}|^{-1/2} |\mathbf{\Sigma}|^{-1/2} \Phi_\rho \left(\mathbf{\Omega}_{diag}^{-1} \boldsymbol{\mu} \right)^{-1} \Phi_\rho \left(\mathbf{V}_{diag}^{-1} \boldsymbol{\mu}_{it}^* \right) \exp\left\{ -\frac{1}{2} \mathbf{Q}_{it} \right\}, \tag{3.4}$$

where $\mathbf{V}^{-1} = \mathbf{\Sigma}^{-1} + \mathbf{\Omega}^{-1}$, $\boldsymbol{\mu}_{it}^* = \mathbf{V}\left(\mathbf{\Omega}^{-1}\boldsymbol{\mu} + \mathbf{\Sigma}^{-1}\mathbf{r}_{it}\right)$, $\mathbf{Q}_{it} = (\mathbf{r}_{it} - \mathbf{r}_{it}^*)\mathbf{W}^{-1}(\mathbf{r}_{it} - \mathbf{r}_{it}^*) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$ such that $\mathbf{W}^{-1} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{\Sigma}^{-1}$, $\mathbf{A} = \mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\left(\mathbf{V} + \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}\mathbf{\Sigma}^{-1}\mathbf{V}\right)\mathbf{\Omega}^{-1}$ and $\mathbf{r}_{it}^* = \mathbf{W}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{\Omega}^{-1}\boldsymbol{\mu}$. Lastly, $\mathbf{M}_{diag}$ denotes the diagonal matrix which contains main diagonal elements of some matrix $\mathbf{M}$.

To obtain (3.4), let $\mathcal{X} = (\mathbf{r}_{it} - \mathbf{u}_{it})' \mathbf{\Sigma}^{-1} (\mathbf{r}_{it} - \mathbf{u}_{it}) + \frac{1}{2}(\mathbf{u}_i - \boldsymbol{\mu})' \mathbf{\Omega}^{-1} (\mathbf{u}_{it} - \boldsymbol{\mu})$ so that the integral in (3.3) becomes $\mathcal{I} = \int_{R_-^d} \exp\left\{-\frac{1}{2}\mathcal{X}\right\} d\mathbf{u}$ with $d = 2$, and it is more convenient to work with the convolution $\mathbf{r} = \mathbf{v} + \mathbf{u}$ when $\mathbf{u} \in \mathbb{R}_-^d$ (i.e., $\mathbf{u}_{it} \leq 0$). Completing the square, we obtain $\mathcal{X} = \mathcal{X}_o + \mathbf{C}$, where $\mathcal{X}_o = (\mathbf{u} - \boldsymbol{\mu}^*)' \mathbf{V}^{-1} (\mathbf{u} - \boldsymbol{\mu}^*)$ and $\mathbf{C} = \mathbf{r}'\mathbf{\Sigma}^{-1}\mathbf{r} + \boldsymbol{\mu}'\mathbf{\Omega}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^{*\prime}\mathbf{V}^{-1}\boldsymbol{\mu}^*$. The latter term, which does not depend on $\mathbf{u}$, can be factorized (like we have shown above) as $\mathbf{C} = (\mathbf{r} - \mathbf{r}^*)\mathbf{W}^{-1}(\mathbf{r} - \mathbf{r}^*) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$. Thus, dealing with $\mathcal{I}$ requires that we compute the integral $\mathcal{I}_o = \int_{R_-^d} \exp\left\{-\frac{1}{2}\mathcal{X}_o\right\} d\mathbf{u}$ by converting it to the integral of the multivariate standard normal distribution. After the standardization transformation, we have that $\mathcal{X}_o = \mathbf{z}'\boldsymbol{\rho}_V^{-1}\mathbf{z}$, where

$z_i = (u_i - \mu_i^*)/\sqrt{V_{ii}}\ \forall\ i = 1, \ldots, d$ and $\boldsymbol{\rho}_V$ is the correlation matrix corresponding to the covariance matrix $\mathbf{V}$ whose main diagonal elements are denoted by V_{ii}. Then, it is straightforward to show that $\mathcal{I}_o = \int_{R_-^d} \exp\left\{-\frac{1}{2}\mathcal{X}_o\right\} d\mathbf{u} = 2\pi \left(\prod_{i=1}^d V_{ii}\right)^{1/2} |\boldsymbol{\rho}_V|^{1/2}\, \Phi_{\boldsymbol{\rho}_V}\left(\cdots, -\mu_i^* V_{ii}^{-1/2}, \cdots\right)$, or $I_o = 2\pi\, |\mathbf{V}|^{1/2}\, \Phi_{\boldsymbol{\rho}_V}\left(\cdots, -\mu_i^* V_{ii}^{-1/2}, \cdots\right)$, where $\Phi_{\boldsymbol{\rho}_V}(b) = (2\pi)^{-d/2}\, |\boldsymbol{\rho}_V|^{-1/2} \int_{\prod_{i=1}^d (-\infty, b_i]} \exp\left\{-\frac{1}{2}\mathbf{z}'\boldsymbol{\rho}_V^{-1}\mathbf{z}\right\} d\mathbf{z}$.

Here, we generalize Pitt & Lee's (1981) results in two important ways. First, we allow for a *general* matrix $\boldsymbol{\Sigma}$. Second, we allow for a multivariate *truncated* normal distribution of $\mathbf{u}_{it}$. With the convention that $\Phi_{\boldsymbol{\rho}}$ denotes the d-variate standard normal distribution function with the correlation matrix $\boldsymbol{\rho} = \left[\omega_{ij}\left(\omega_{ii}\omega_{jj}\right)^{-1/2};\ i, j = 1, \ldots, d\right]$, the formula in (3.4) applies to the general d-dimensional multivariate case.

3.2 Model II: Hierarchical Dependence

In contrast to Model I which imposes no structure onto the order of causality in the dependence between cost efficiency and market power, Model II postulates an *a priori* hierarchical dependence between $u_{1,it}$ and $u_{2,it}$, whereby the former follows a one-sided distribution conditional on the latent market power which, in turn, also follows a one-sided distribution. More specifically, we make the following stochastic assumptions for Model II:

$$\begin{bmatrix} v_{1,it} \\ v_{2,it} \end{bmatrix} \sim \mathbb{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{3.5a}$$

$$u_{1,it}|u_{2,it} \sim \mathbb{N}^+\left(\mathbf{z}_{1,it}'\boldsymbol{\gamma}_1 + \psi_1 u_{2,it},\ \omega_1^2\right), \tag{3.5b}$$

$$u_{2,it} \sim \mathbb{N}^+\left(\mathbf{z}_{2,it}'\boldsymbol{\gamma}_2,\ \omega_2^2\right)\ . \tag{3.5c}$$

In this model, the latent market power is dependent on a vector of contextual covariates $\mathbf{z}_{2,it}$ and follows a truncated normal distribution to ensure its non-negativity. The latent cost inefficiency, which depends on a vector of contextual covariates $\mathbf{z}_{1,it}$, follows a truncated normal distribution and, importantly, is also affected by the latent market power through its mean.

We note that the specification of the hierarchical dependence between $u_{1,it}$ and $u_{2,it}$ in (3.5) is different from that implied by the bivariate truncated normal distribution employed in Model I. To see this clearly, recognize that, if $\mathbf{u}_{it} \sim \mathbb{N}^+\left(\left[(\mathbf{z}_{1,it}'\boldsymbol{\gamma}_1)', (\mathbf{z}_{2,it}'\boldsymbol{\gamma}_2)'\right]', \boldsymbol{\Omega}\right)$, then it follows that

$$u_{1,it}|u_{2,it} \sim \mathbb{N}^+\left(\mathbf{z}_{1,it}'\boldsymbol{\gamma}_1 + \frac{\omega_{12}}{\omega_{22}}\left(u_{2,it} - \mathbf{z}_{2,it}'\boldsymbol{\gamma}_2\right),\ \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}\right) \tag{3.6a}$$

$$u_{2,it}|u_{1,it} \sim \mathbb{N}^+\left(\mathbf{z}_{2,it}'\boldsymbol{\gamma}_2 + \frac{\omega_{12}}{\omega_{11}}\left(u_{1,it} - \mathbf{z}_{1,it}'\boldsymbol{\gamma}_1\right),\ \omega_{22} - \frac{\omega_{12}^2}{\omega_{11}}\right), \tag{3.6b}$$

implying that the two variables are dependent through the covariance parameter ω_{12}. Therefore, the hierarchical dependence is different from the dependence implied by the joint specification.

3.3 Model III: Dependence via Conditional Distributions

Model III relaxes Model II further by also allowing for the dependence of the latent market power on the cost efficiency via its own conditional distribution. Thus, the cross-equation dependence is now explicitly bidirectional. That is, we assume that

$$\begin{bmatrix} v_{1,it} \\ v_{2,it} \end{bmatrix} \sim \mathbb{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{3.7a}$$

$$u_{1,it}|u_{2,it} \sim \mathbb{N}^{+}\left(\mathbf{z}'_{1,it}\boldsymbol{\gamma}_1 + \psi_1 u_{2,it},\ \omega_1^2\right), \tag{3.7b}$$

$$u_{2,it}|u_{1,it} \sim \mathbb{N}^{+}\left(\mathbf{z}'_{2,it}\boldsymbol{\gamma}_2 + \psi_2 u_{1,it},\ \omega_2^2\right). \tag{3.7c}$$

Clearly, this model (like the previous two) does not *a priori* restrict signs of the parameters controlling the dependence between u_1 and u_2 in the two conditional distributions:

$$p\left(u_{1,it}|u_{2,it},\cdot\right) = \left(2\pi\omega_1^2\right)^{-1/2}\exp\left\{-\frac{1}{2\omega_1^2}\left(u_{1,it} - \mathbf{z}'_{1,it}\boldsymbol{\gamma}_1 - \psi_1 u_{2,it}\right)^2\right\}\times$$
$$\Phi\left(\left(\mathbf{z}'_{1,it}\boldsymbol{\gamma}_1 + \psi_1 u_{2,it}\right)/\omega_1\right)^{-1}\mathbb{1}\left(u_{1,it}\geq 0\right) \tag{3.8}$$

and

$$p\left(u_{2,it}|u_{1,it},\cdot\right) = \left(2\pi\omega_2^2\right)^{-1/2}\exp\left\{-\frac{1}{2\omega_2^2}\left(u_{i2} - \mathbf{z}'_{2,it}\boldsymbol{\gamma}_2 - \psi_2 u_{1,it}\right)^2\right\}\times$$
$$\Phi\left(\left(\mathbf{z}'_{2,it}\boldsymbol{\gamma}_2 + \psi_2 u_{1,it}\right)/\omega_2\right)^{-1}\mathbb{1}\left(u_{2,it}\geq 0\right). \tag{3.9}$$

The joint distribution $p\left(u_{1,it},u_{2,it}|\cdot\right) = p\left(u_{1,it}|u_{2,it},\cdot\right)p\left(u_{2,it}|\cdot\right) = p\left(u_{2,it}|u_{1,it},\cdot\right)p\left(u_{1,it}|\cdot\right)$ is unavailable given that the marginal distributions are not specified. Combining with the relevant terms in the posterior kernel distribution, we have

$$p\left(u_{1,it},u_{2,it}|\boldsymbol{\Xi},\boldsymbol{\theta}\right) = \exp\left\{-\frac{1}{2}\left(\mathbf{u}_{it}-\mathbf{r}_{it}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{u}_{it}-\mathbf{r}_{it}\right)\right\}p\left(u_{1,it},u_{2,it}|\cdot\right), \tag{3.10}$$

where $\boldsymbol{\Xi}$ denotes the available data.

4 Empirical Application

We showcase our proposed unified system-based model by applying it to study the interplay between monopolistic power and cost efficiency in the U.S. commercial banking industry.

4.1 Data

The annual bank-level year-end data that we use in this paper come from Koetter et al. (2012) and originate in Call Reports of the Federal Reserve System. The sample includes all FDIC-insured commercial banks with available data between 1984 and 2007. We exclude banks reporting negative values for assets, equity, outputs and prices. Following Stiroh & Strahan (2003) and Koetter et al. (2012), we also exclude banks in the District of Columbia and South Dakota due to their exceptional laws concerning the credit card business practices. To mitigate the influence of outliers, we also truncate input prices at the 1st and 99th percentiles of their respective empirical distributions. All nominal quantities are deflated using the 2005 Consumer Price Index for all urban consumption published by the U.S. Bureau of Labor Statistics. The operational dataset is an unbalanced panel of 17,148 banks with a total of 159,061 observations.

We model the bank's production technology using the commonly used "intermediation approach" of Sealey & Lindley (1977), according to which a bank's balance sheet is assumed to capture the essential structure of its core business. Liabilities, together with physical capital and labor, are taken as inputs to the bank's production process, whereas assets (other than physical) are considered as outputs. Specifically, we define two output variables: securities (y_1) and loans (y_2). The inputs are fixed assets (x_1), labor (x_2) and borrowed funds (x_3). Total costs (C) equals

Table 1. Data Summary Statistics

Variable	Mean	5th Perc.	Median	95th Perc.	Units of Measurement
			Production Variables		
Securities (y_1)	102,384.0	2,806.5	21,011.2	204,704.0	'000 real 2005 USD
Loans (y_2)	332,621.4	7,459.1	45,115.0	567,837.6	'000 real 2005 USD
Price of Fixed Assets (w_1)	36.24	11.57	27.19	94.44	% pt.
Price of Labor (w_2)	33.52	18.54	31.00	57.10	'000 real 2005 USD
Price of Borrowed Funds (w_3)	4.22	1.55	4.03	7.09	% pt.
Equity (k)	43,090.1	1,431.0	7,388.9	74,154.6	'000 real 2005 USD
Cost (C)	28,953.5	1,022.1	4,811.3	48,334.2	'000 real 2005 USD
Revenue (R)	45,766.0	1,438.5	6,834.4	72,418.2	'000 real 2005 USD
			Contextual Variables		
Assets	530,147.9	16,843.2	81,443.3	874,989.5	'000 real 2005 USD
Top Hundred by Asset Size	0.01				unit-free
Asset Market Share in the State	0.51	0.01	0.10	1.34	% pt.
# of Mergers in the State	20.66	0.00	13.00	61.00	cardinal number
Equity-to-Assets Ratio	9.29	5.75	8.66	14.97	% pt.
Securities-to-Assets Ratio	28.10	6.66	26.45	55.04	% pt.
HHI for Loans	0.45	0.27	0.41	0.77	% pt.
Non-interest Income Share	9.09	2.58	7.71	19.74	% pt.
Loan-Loss Provision Share	0.73	0.00	0.32	2.84	% pt.
Loan-Loss Reserve Share	1.57	0.66	1.31	3.32	% pt.
Z-Score	48.01	5.77	34.24	132.60	unit-free
State Unemployment Rate	5.35	2.97	5.12	8.46	% pt.
State DPI	168,127.5	21,181.0	109,528.5	522,984.7	'000 real 2005 USD

the sum of expenses on these inputs. Input prices (w_1, w_2, w_3) are obtained by dividing expenses on these items by their respective quantities. We also include equity capital (k) as an additional input to the production technology. The treatment of equity as an input to banking production technology is usually motivated by the argument that banks may use the latter as a source of loanable funds and thus as a cushion against losses. Due to the unavailability of the price of equity, we follow Berger & Mester (1997, 2003) in modeling k as a quasi-fixed input. Total revenue (R) is the sum of revenues from the two output categories.

In addition to the production-related covariates described above, we also incorporate a number of variables capturing various bank's characteristics, both internal and external, in our analysis. These variables are intended to contextualize the economic environment in which banks operate as well as to control for their different business strategies related to efficiency and market power. Specifically, we let the cost efficiency and Lerner index be functions of contextual variables that capture heterogeneity across banks along various dimensions including the scope and overall competitiveness of the market, the bank's size and product mix as well as risk taking. By controlling for banks' heterogeneous features, we seek to more "cleanly" isolate the interplay between cost efficiency and monopolistic power. In the language of econometric models described in Section 3, such contextual control covariates are the "$\mathbf{z}$" variables affecting the conditional time-varying bank-specific means of two one-sided latent variables related to cost inefficiency and the Lerner index. These variables are as follows: (i) the bank's total assets capture its size and scale of operations; (ii) an indicator for the top-hundred banks (by the asset size) in a given year and (iii) the bank's asset market share in a given state are included as observable measures of the bank's dominance in the market (see Stiroh & Strahan, 2003; Boyd & De Nicolo, 2005; Hannan & Prager, 2004, 2009); (iv) the number of bank mergers in the state in a given year also captures competitive pressures in the

Table 2. Posterior Mean Estimates of Scale Elasticity

Model	Mean Estimate	95% Bayes Interval
(I)	0.8677	(0.8014; 0.9336)
(II)	0.9083	(0.8147; 1.0010)
(III)	0.8359	(0.7581; 0.9136)
(IV)	0.8858	(0.8220; 0.9496)

market; (v) the bank's ratio of equity to total assets measuring its capitalization is meant to control for factors contributing to bank distress (e.g., Gan, 2004); (vi) the bank's ratio of securities to total assets, (vii) the Hirschman-Herfindahl index across the banks' different types of loans and (viii) the share of non-interest income are all controlling for the possibility that greater competition might entice banks to engage in nontraditional activities as well as to more actively seek diversification of their portfolio (Koetter et al., 2012); (ix) the share of loan-loss provisions and (x) loan-loss reserves in the bank's total loans proxy for the credit risk, whereas (xi) the bank's z-score[7] proxies for the overall risk of bank failure; (xii) the disposable personal income and (xiii) the unemployment rate in the state are the controls for macroeconomic conditions which may affect the competition (Chirinko & Fazzari, 2000) as well as efficiency. We also include three indicator variables reflective of institutional changes in states that correspond to deregulation in the intrastate branching (by means of mergers and acquisitions), the interstate expansion (via bistate agreements) and the post-IBBEA interstate banking. The chosen contextual variables are all likely to greatly influence bank's business strategies in pursuit of the maximum franchise value (Demsetz, Saidenberg & Strahan, 1996; DeYoung & Rice, 2004) with important implications for its efficiency and/or market power. See Table 1 of Koetter et al. (2012) for details on the construction and rationale behind the variables. Table 1 above presents the data summary statistics.

4.2 Results

In this section, we report the results from our unified system-based model in (2.6) estimated using the three specifications described in Section 3. These econometric models are respectively referred to as the Model I, II and III. In all three cases, we assume the translog cost function and allow for non-neutral temporal shifts in the bank's cost frontier as well as let the means of latent u_C and u_L be time-varying and bank-specific conditional on the contextual variables capturing heterogeneous bank characteristics. In line with the intuition outlined earlier, we include the contextual variables summarized in Section 4.1 in the covariate set of the mean function of cost inefficiency u_C and/or the log-Lerner-index function u_L; both means also include the time trend and its square. (For complete variable lists corresponding to each mean function, see Table 4). The bank's cost efficiency is computed as $\exp\{-\widehat{u}_C\}$, whereas the (automatically adjusted) Lerner index is recovered as $1 - \exp\{-\widehat{u}_L\}$.

To highlight the merits of our proposed system-based approach, we also examine the relationship between market power and cost efficiency employing the most popular strategy in the literature whereby we estimate the stochastic cost frontier in (2.6a) alone without accounting for the joint dependence of the Lerner index and efficiency. Following Koetter et al. (2012) and many others, the fitted cost frontier is then used along with raw data on total revenues to construct the "adjusted" Lerner index as follows: $\widehat{L} = \big(R - \widehat{C} \times \sum_m \widehat{\varepsilon}_{y_m}\big)/R$, where both $\widehat{C}$ and $\widehat{\varepsilon}_{y_m}$ are obtained from the estimated stochastic cost frontier. The relation between the cost efficiency estimates and the

[7]Computed following Laeven & Levine (2009) and using the four-year rolling-window standard deviations.

Table 3. Posterior Mean Estimates of Cost Efficiency and Market Power

Model	Mean Estimate	95% Bayes Interval
	Model I	
Efficiency	0.8230	(0.7928; 0.8534)
Lerner Index	0.3349	(0.2885; 0.3812)
	Model II	
Efficiency	0.8090	(0.7530; 0.8649)
Lerner Index	0.3321	(0.2737; 0.3904)
	Model III	
Efficiency	0.7983	(0.7288; 0.8675)
Lerner Index	0.3589	(0.3015; 0.4165)
	Model IV	
Efficiency	0.7675	(0.6704; 0.8644)
Lerner Index	0.4356	(0.3858; 0.4854)

Lerner index is then analyzed in the second stage. We refer to this model hereinafter as Model IV. It is primarily meant to illustrate the empirical sensitivity of the findings to proper modeling of the simultaneous determination of both the bank's efficiency and market power.

Before we proceed to the discussion of the main results concerning the market power – efficiency nexus and its accompanying hypotheses, we first examine the estimates of the scale elasticity. The posterior mean estimates of ε_y along with their 95% credible interval from the four estimated models are reported in Table 2. The estimates are of interest because they gauge returns to scale in the industry. Specifically, the bank is said to exhibit increasing/constant/decreasing returns to scale if the scale elasticity (of cost) is less than/equal to/greater than one. While the posterior mean (point) estimates from all four models suggest that, on average, banks operate at increasing returns to scale during our sample period, in the case of Model II the 95% posterior coverage region includes unity thereby suggesting roughly constant returns to scale. All other models however indicate that the banking industry exhibits significant scale economies, consistent with the recent findings (e.g., Wheelock & Wilson, 2012; Hughes & Mester, 2013; Malikov, Restrepo-Tobón & Kumbhakar, 2015).

Table 3 presents the estimates of primary interest to our paper. The reported are the posterior mean estimates of the bank's cost efficiency and the Lerner index from the four models over the entire sample. Among the three specifications of our system-based approach, Model I yields the highest estimates of the mean cost efficiency for banks at around 0.82, while Model III produces the lowest estimates that are, on average, about 2.5 basis points lower. Interestingly, when we estimate the cost frontier without any accommodation of the joint dependence of the bank's market power and efficiency (Model IV), we obtain efficiency scores that are even lower with the average posterior estimate of 0.77. The differences in results from the two types of models (our preferred system-based estimator vs. a more popular single-equation specification) are more evident when we contrast their estimates of the Lerner index. Model IV appears to over-estimate the monopolistic power exercised by the banks in our sample, with the pooled mean posterior estimate being as high as 0.44 versus the value of the corresponding statistic from our system-based Models I–III ranging between 0.33 and 0.36.

We are now ready to formally examine the relationship between the bank's market power and cost efficiency, the focal point of our paper. In Table 4, we report the posterior mean estimates (along with their 95% Bayes intervals) of the parameters describing the distribution(s) of the two latent variables of interest: cost inefficiency u_C and the log-Lerner-index function u_L. Our proposed

Table 4. Posterior Mean Estimates of Parameters across Models I–IV

	(I)	(II)	(III)	(IV)
		Mean of Cost Inefficiency		
Constant	–1.335	–0.717	–0.445	–0.373
	(–1.515; –0.877)	(–0.944; –0.342)	(–0.525; –0.316)	(–0.551; –0.212)
t	0.035	0.042	0.017	0.005
	(0.017; 0.044)	(0.030; 0.052)	(0.012; 0.023)	(0.002; 0.013)
t^2	–0.004	0.003	0.001	0.001
	(–0.006; –0.002)	(–0.004; 0.005)	(–0.002; 0.003)	(–0.002; 0.003)
log(Size)	0.377	0.173	0.216	0.105
	(0.221; 0.414)	(0.005; 0.224)	(0.130; 0.245)	(0.007; 0.155)
Equity-to-Assets Ratio	0.251	0.180	0.313	0.216
	(0.181; 0.322)	(0.103; 0.199)	(0.277; 0.366)	(0.189; 0.256)
Loan-Loss Provision Share	0.414	0.255	0.202	0.104
	(0.303; 0.525)	(0.188; 0.344)	(0.177; 0.287)	(0.005; 0.153)
Loan-Loss Reserve Share	0.382	0.214	0.288	0.105
	(0.301; 0.405)	(0.187; 0.289)	(0.217; 0.322)	(0.007; 0.114)
Z-Score	–0.155	0.224	–0.188	0.150
	(–0.182; –0.051)	(0.117; 0.344)	(–0.226; –0.032)	(0.008; 0.178)
log(DPI)	0.036	0.043	0.025	–0.015
	(0.017; 0.045)	(0.025; 0.062)	(0.010; 0.041)	(–0.022; –0.006)
Unemployment Rate	–0.040	–0.035	–0.044	–0.017
	(–0.051; –0.032)	(–0.044; –0.022)	(–0.051; –0.030)	(–0.001; –0.024)
Intrastate Deregulation	0.355	0.132	0.177	0.044
	(0.228; 0.381)	(0.065; 0.187)	(0.045; 0.193)	(0.035; 0.057)
Interstate Deregulation	–0.312	–0.044	–0.225	–0.104
	(–0.420; –0.255)	(–0.055; –0.038)	(–0.326; –0.189	(–0.005; –0.176)
IBBEA Deregulation	0.286	0.188	0.142	0.176
	(0.133; 0.351)	(0.102; 0.197)	(0.133; 0.171)	(0.103; 0.222)
u_2		0.166	0.182	
		(0.103; 0.177)	(0.144; 0.202)	

(continued on the next page)

NOTE: The 95% credible intervals in parentheses.

system-based approach to modeling joint dependence of the latent market power and cost efficiency presents a natural tool to statistically assess the relationship and to formally discriminate between the two competing hypotheses: QLH versus ESH. More specifically, depending on the econometric formulation of our system, we can test the sign of the relationship by examining *(i)* the covariance between u_C and u_L in Model I, *(ii)* the coefficient of u_L appearing in the mean function of u_C in Model II or *(iii)* the coefficients of u_L and u_C respectively appearing in the mean functions of u_C and u_L in Model III. From Table 4, we see that, across all three unified system-based Models I–III, the relevant parameters regulating the dependence between u_L and u_C are significantly positive. Since u_C is a decreasing function of the cost efficiency while u_L is an increasing function of the market power, the data thus lend support to the QLH whereby the greater monopolistic power generally permits banks to operate at lower efficiency levels.[8] This is consistent with earlier findings by Koetter & Vins (2008), Delis & Tsionas (2009), Turk Ariss (2010) and Dong et al. (2016) for banks in the U.S. and other countries. Also, recall that our result is conditional on heterogeneous bank characteristics which we control for in the estimation of means of u_C and u_L. Further, the

[8]Since a *positive* relationship between u_C and u_L imply a *negative* relationship between cost efficiency ($\exp\{-u_C\}$) and the Lerner measure of market power ($1 - \exp\{-u_L\}$).

Table 4. Posterior Mean Estimates of Parameters across Models I–IV (cont.)

	(I)	(II)	(III)	(IV)
		Mean of the log Lerner Index		
Constant	–0.355	–0.714	–0.525	
	(–0.617; –0.188)	(–1.414; –0.353)	(–0.871; –0.212)	
t	0.044	0.032	0.030	
	(0.015; 0.055)	(0.020; 0.048)	(0.017; 0.055)	
t^2	–0.004	–0.002	–0.001	
	(–0.008; –0.002)	(–0.001; –0.003)	(–0.012; 0.120)	
log(Size)	0.366	0.289	0.312	
	(0.216; 0.457)	(0.155; 0.317)	(0.181; 0.416)	
Top Hundred Bank	0.422	0.388	0.455	
	(0.289; 0.588)	(0.101; 0.560)	(0.317; 0.588)	
Asset Market Share in State	0.203	0.181	0.225	
	(0.113; 0.352)	(0.044; 0.203)	(0.141; 0.327)	
# Mergers in State	0.456	0.188	0.417	
	(0.382; 0.551)	(0.072; 0.277)	(0.226; 0.617)	
Securities Share	0.388	0.217	0.181	
	(0.144; 0.524)	(0.044; 0.513)	(0.072; 0.226)	
HHI for Loans	0.727	0.515	0.103	
	(0.551; 0.827)	(0.313; 0.688)	(0.064; 0.203)	
Non-interest Income Share	0.332	0.202	0.188	
	(0.187; 0.482)	(0.103; 0.355)	(0.065; 0.254)	
log(DPI)	0.316	0.217	–0.016	
	(0.188; 0.415)	(0.188; 0.335)	(–0.035; 0.044)	
Unemployment Rate	0.188	0.103	–0.022	
	(0.055; 0.213)	(0.054; 0.210)	(–0.036; –0.015)	
Intrastate Deregulation	0.316	–0.117	–0.025	
	(0.188; 0.440)	(–0.221; –0.073)	(–0.033; 0.015)	
Interastate Deregulation	0.283	0.132	–0.017	
	(0.155; 0.318)	(–0.071; 0.218)	(–0.022; 0.030)	
IBBEA Deregulation	0.187	0.351	–0.032	
	(0.133; 0.220)	(0.144; 0.447)	(–0.052; 0.071)	
u_1			0.228	
			(0.170; 0.316)	
		Variance–Covariance		
var(u_1)	0.317	0.225	0.228	0.187
	(0.285; 0.388)	(0.188; 0.287)	(0.186; 0.303)	(0.115; 0.214)
var(u_2)	0.455	0.317	0.285	
	(0.388; 0.515)	(0.289; 0.355)	(0.212; 0.317)	
cov(u_1, u_2)	0.224			
	(0.216; 0.277)			
Log Bayes Factor	22.351	1.000	2.305	

NOTE: The 95% credible intervals in parentheses. Bayes Factors are relative to Model II.

differences in our formulation of the dependence between the cost efficiency and the Lerner index across specifications I through III also allow us to implicitly assess the underlying nature of the relationship between the two. Concretely, the log Bayes factors[9] reported at the bottom of Table 4 indicate that our data distinctly favor the "agnostic" Model I over the two alternative econometric specifications of the joint dependence between u_C and u_L. Thus, the selected model implies that the data are *not* revealing of a clear causal directionality in the relationship between the bank's efficiency and monopolistic power.

The QLH-consistent negative relationship between the bank's cost efficiency and market power can be vividly seen by looking at Figure 1. The top row of Figure 1 depicts contours of bivariate kernel densities[10] of the bank-level estimates of cost efficiency and the Lerner index for all three specification of our proposed model. These plots allow us to assess the relationship not just at a given moment (say, average) but distribution-wise. Conversely, plots in the bottom row of Figure 1 enable us to examine the relationship at different quantiles. Specifically, they show the estimated 10th, 25th, 50th, 75th and 90th quantiles of the bank's cost efficiency conditional on its Lerner index. The fitted conditional quantile functions are obtained by inverting the nonparametrically estimated kernel conditional cdf of the cost efficiency given the market power.[11] To avoid any confusion, we would like to stress that the plotted are not the confidence bounds. Both kinds of plots suggest that, when the simultaneous determination of the banks' efficiency and the market power is modeled explicitly (as in Models I–III), the two exhibit a strong negative relationship. To the contrary of the above results from our system-based models, a single-equation Model IV points to a positive relationship between the bank's cost efficiency and the Lerner index which is in line with the ESH. To see this, consider Figure 2 which plots the bivariate kernel density and the conditional quantile functions for the estimates from Model IV. For instance, Koetter et al. (2012) also find such a positive relationship using the method like the one of Model IV. However, since Model IV does not explicitly formulate the joint dependence between the market power and efficiency, we do not have a direct way to formally test for the sign of the relationship between the two. While one would normally be tempted to run a second-stage regression as widely done in the literature, the latter procedure would not produce valid inference in the light of the problems discussed in the Introduction. Thus, any inference, even informal, on the basis of patterns discernible from Figure 2 is likely to be misleading, especially because the estimates of both measures are prone to simultaneity and misspecification biases due to Model IV's failure to meaningfully accommodate the joint dependence of the two.

We conclude by briefly commenting on the results pertaining to contextual controls. Most bank characteristics have significant effects on the mean cost inefficiency and marker power, and the effects are largely of expected signs with very few reversals across model specifications. We find that, over time, banks have generally acquired more market power while having become more cost *in*efficient. Rather expectedly, the results from all models also suggest that larger banks tend to be less efficient but also to exert greater monopolistic power in the markets than their smaller counterparts. Consistent with one's intuition, the largest banks as well as, more generally, banks with larger market share are estimated to have more market power. Same holds for the institutions operating in less competitive markets as proxied by the number of mergers at the state level. The banks with higher market power are also found to be those with less diversified loan portfolios (higher HHI for loans) and those more heavily engaged in nontraditional activities. Based on the results from Model I most preferred by the data, we also find that the deregulation appears to have

[9] Computed using a Laplace approximation (DiCiccio, Kass, Raftery & Wasserman, 1997).

[10] We use an axis-aligned bivariate Gaussian kernel, evaluated on a square grid using the normal reference bandwidth.

[11] We employ the second-order Gaussian kernel and the cross-validated bandwidth.

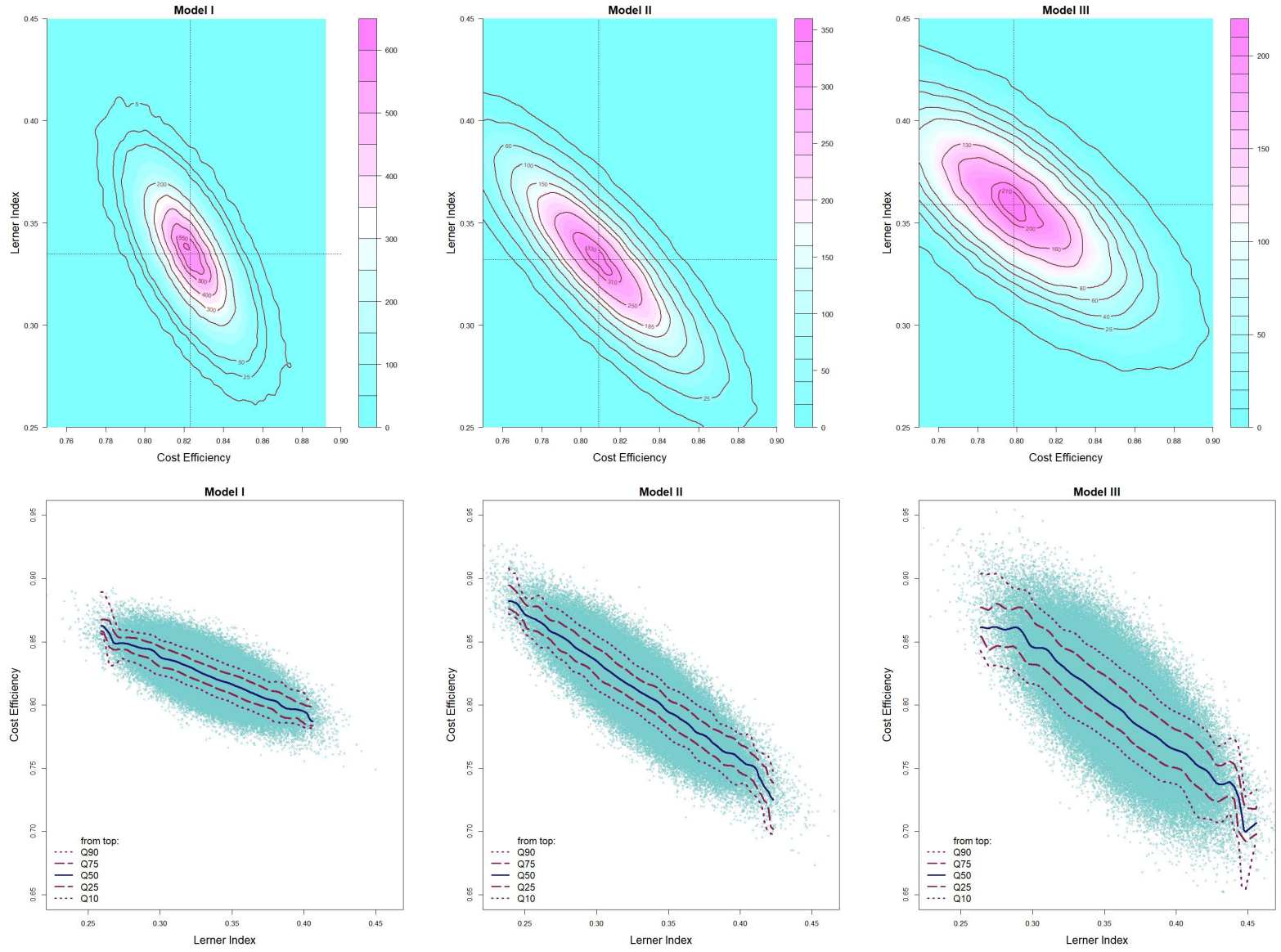

Figure 1. Banks' Cost Efficiency and Market Power (Models I–III):
Bivariate Kernel Densities (top row) and Nonparametric Quantile Regressions (bottom tow)

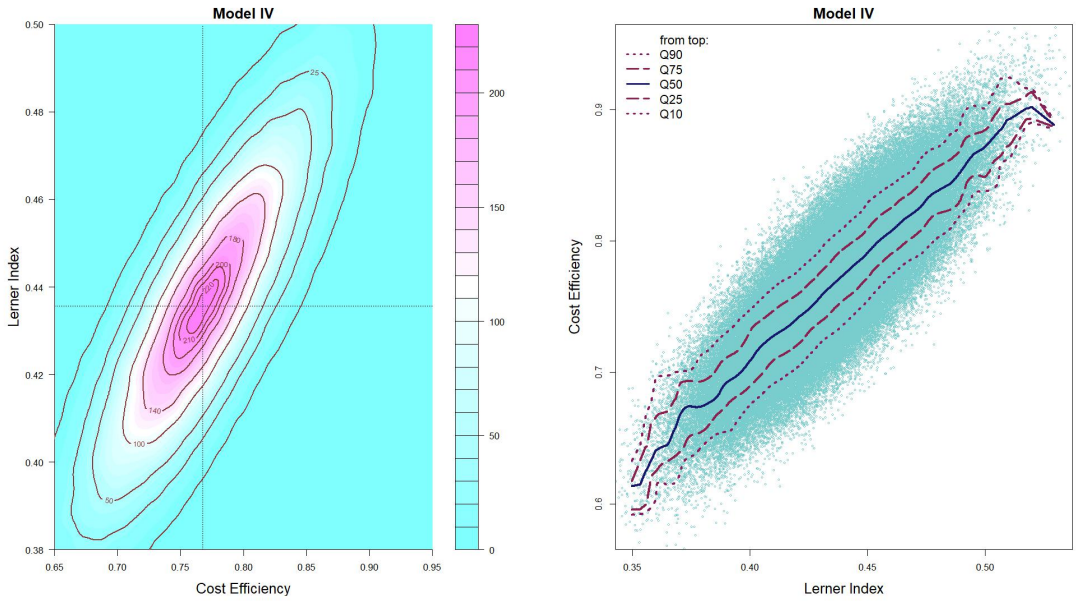

Figure 2. Bivariate Kernel Density and Nonparametric Quantile Regression of Banks' Cost Efficiency and Market Power (Model IV)

positively contributed to the monopolization of the industry. The model also suggests that banks with lower z-scores (higher probability of insolvency) tend to be more cost inefficient.

5 Conclusion

This paper develops a novel unified econometric methodology for the formal examination of the market power – cost efficiency nexus. Our approach can meaningfully accommodate a mutually dependent relationship between the firm's cost efficiency and marker power (as measured by the Lerner index) by explicitly modeling the simultaneous/endogenous determination of the two in a system of nonlinear equations consisting of the firm's cost frontier and the revenue-to-cost ratio equation derived from its stochastic revenue function. Both the firm's cost efficiency and marker power index are estimated jointly and derived from a single unified model thus enabling us to interpret and analyze them on a common ground. Our framework places no *a priori* restrictions on the sign of the dependence between the firm's monopolistic power and efficiency as well as allows for different hierarchical orderings between the two, enabling us to meaningfully discriminate between competing quiet life and efficient structure hypotheses. Among other benefits, our approach completely obviates the need for second-stage regressions of the cost efficiency estimates on the constructed market power measures which, while widely prevalent in the literature, suffer from multiple econometric problems as well as lack internal consistency/validity.

We showcase our methodology by applying it to a panel of U.S. commercial banks in 1984–2007. To draw statistical inference, we consider three alternative econometric specifications of our unified system-based model which we estimate using MCMC methods. Regardless of the econometric specification of the model used, the data consistently point to a negative dependence between the bank's cost efficiency and the Lerner index thus providing empirical evidence in support of the quiet life hypothesis. This finding is reversed when we employ a traditional two-stage analysis where the cost efficiency and the adjusted Lerner index are estimated separately without allowing

for a simultaneous determination of the two. The latter highlights the pivotal importance of a proper econometric modeling of the market power – efficiency relationship which a popular two-stage analysis is unable to deliver.

Appendix: MCMC Techniques

Across all models (where appropriate), we use the following priors for $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, ψ_1, ψ_1, $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$: $\gamma_j \sim \mathbb{N}(\mathbf{0}, 10^4\mathbf{I})$ for $j = 1, 2$; $\psi_j \sim \mathbb{N}(0, 10^4)$ independently for $j = 1, 2$; $p(\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{(-\bar{n}+1)} \exp\{tr[\bar{\mathbf{A}}\boldsymbol{\Omega}^{-1}]\}$ and $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{(-\bar{n}+1)} \exp\{tr[\bar{\mathbf{A}}\boldsymbol{\Sigma}^{-1}]\}$, both from the inverse-Wishart distribution with $\bar{n} = 1$ and $\bar{\mathbf{A}} = 10^{-4}\mathbf{I}$.

A Model I

The augmented kernel posterior distribution is

$$p\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \mathbf{u}|\boldsymbol{\Xi}\right) \propto |\boldsymbol{\Omega}|^{-nT/2} |\boldsymbol{\Sigma}|^{-nT/2} \Phi_\rho \left(\boldsymbol{\Omega}_{diag}^{-1}\boldsymbol{\mu}\right)^{-nT} \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[(\mathbf{r}_{it} - \mathbf{u}_{it})'\boldsymbol{\Sigma}^{-1}(\mathbf{r}_{it} - \mathbf{u}_{it}) - (\mathbf{u}_{it} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{u}_{it} - \boldsymbol{\mu})\right]\right\}, \tag{A.1}$$

where $\mathbf{r}_{it} = \mathbf{y}_{it} - \begin{bmatrix} \mathbf{x}_{it}'\boldsymbol{\beta} \\ f(\mathbf{x}_{it}; \boldsymbol{\beta}) \end{bmatrix}$.

Like we have showed in Section 3, while it is certainly possible to integrate the latent variables out, the resulting posterior is however highly nonlinear. Since $\boldsymbol{\beta}$ enters the second equation in (3.1) in a nonlinear way, we need to construct an efficient proposal distribution to use with the Metropolis-Hastings algorithm.

Let $\widehat{\boldsymbol{\beta}}$ be some estimator of $\boldsymbol{\beta}$, say, the least squares estimator applied to the cost function in (3.1a). Then, the scale elasticity can then be computed as $\sum_m \widehat{\varepsilon}_{y_{m,it}} = \mathbf{x}_{o,it}'\mathbf{R}\widehat{\boldsymbol{\beta}}$. Linearizing $f(\mathbf{x}_{it}; \boldsymbol{\beta}) \equiv \ln\left(\sum_m \varepsilon_{y_{m,it}}\right)$, we obtain that $\widehat{f}_{it}(\mathbf{x}; \boldsymbol{\beta}) \simeq \ln\left(\sum_m \widehat{\varepsilon}_{y_{m,it}}\right) - 1 + \left(\sum_m \widehat{\varepsilon}_{y_{m,it}}\right)^{-1} \mathbf{x}_{o,it}'\mathbf{R}\boldsymbol{\beta}$. We next rewrite the system in (3.1) as follows:

$$y_{1,it} - u_{1,it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{1,it} \tag{A.2a}$$

$$y_{2,it} - \ln\left(\mathbf{x}_{o,it}'\mathbf{R}\widehat{\boldsymbol{\beta}}\right) + 1 - u_{2,it} \simeq \left(\mathbf{R}'\mathbf{x}_{o,it}\left(\mathbf{x}_{o,it}'\mathbf{R}\widehat{\boldsymbol{\beta}}\right)^{-1}\right)'\boldsymbol{\beta} + v_{2,it} \equiv \widetilde{\mathbf{x}}_{o,it}'\boldsymbol{\beta} + v_{2,it}. \tag{A.2b}$$

For known values of $\mathbf{u}_{it}$, the GLS estimator of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}^* = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{X}_{it}\boldsymbol{\Sigma}^{-1}\mathbf{X}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{X}_{it}\boldsymbol{\Sigma}^{-1}\mathbf{Y}_{it}, \tag{A.3}$$

where we let $\mathbf{Y}_{it} = \begin{bmatrix} y_{1,it} - u_{1,it} \\ y_{2,it} - \ln\left(\mathbf{x}_{o,it}'\mathbf{R}\widehat{\boldsymbol{\beta}}\right) + 1 - u_{2,it} \end{bmatrix}$ and $\mathbf{X}_{it} = \begin{bmatrix} \mathbf{x}_{it} \\ \widetilde{\mathbf{x}}_{o,it} \end{bmatrix}$. The corresponding GLS variance-covariance matrix is $\mathbf{V}^* = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{X}_{it}\boldsymbol{\Sigma}^{-1}\mathbf{X}_{it}'\right)^{-1}$.

The proposal distribution is $\mathbb{N}_k(\boldsymbol{\beta}^*, h\mathbf{V}^*)$, where $h > 0$ is a certain constant. If we draw a candidate $\boldsymbol{\beta}^c \sim \mathbb{N}_k(\boldsymbol{\beta}^*, h\mathbf{V}^*)$ and the chain is currently at $\boldsymbol{\beta}^o$, according to the independence Metropolis-Hastings proposal, the acceptance probability is

$$\min\left\{1,\ \frac{\exp\left\{-\frac{1}{2}tr\left[\mathbf{Q}(\boldsymbol{\beta}^c)\boldsymbol{\Sigma}^{-1}\right] - \frac{1}{2h^2}\left(\boldsymbol{\beta}^c - \boldsymbol{\beta}^*\right)'\mathbf{V}^{*-1}\left(\boldsymbol{\beta}^c - \boldsymbol{\beta}^*\right)\right\}}{\exp\left\{-\frac{1}{2}tr\left[\mathbf{Q}(\boldsymbol{\beta}^o)\boldsymbol{\Sigma}^{-1}\right] - \frac{1}{2h^2}\left(\boldsymbol{\beta}^o - \boldsymbol{\beta}^*\right)'\mathbf{V}^{*-1}\left(\boldsymbol{\beta}^o - \boldsymbol{\beta}^*\right)\right\}}\right\}, \tag{A.4}$$

where $\mathbf{Q}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{y}_{it} - \begin{bmatrix}\mathbf{x}_{it}'\boldsymbol{\beta} \\ f(\mathbf{x}_{it};\boldsymbol{\beta})\end{bmatrix}\right)\left(\mathbf{y}_{it} - \begin{bmatrix}\mathbf{x}_{it}'\boldsymbol{\beta} \\ f(\mathbf{x}_{it};\boldsymbol{\beta})\end{bmatrix}\right)'$.

An alternative is to use a random walk Metropolis-Hastings proposal in which $\boldsymbol{\beta}^c \sim \mathbb{N}_k(\boldsymbol{\beta}^o, h\mathbf{V}^*)$. The acceptance probability then becomes

$$\min\left\{1,\ \frac{\exp\left\{-\frac{1}{2}tr\left[\mathbf{Q}(\boldsymbol{\beta}^c)\boldsymbol{\Sigma}^{-1}\right]\right\}}{\exp\left\{-\frac{1}{2}tr\left[\mathbf{Q}(\boldsymbol{\beta}^o)\boldsymbol{\Sigma}^{-1}\right]\right\}}\right\}. \tag{A.5}$$

Also, note that the proposal distributions can be constructed using the direct least squares estimator of the cost function in (A.2a). In this case, $\boldsymbol{\beta}^*$ and $\mathbf{V}^*$ in the above discussion is to be replaced with $\boldsymbol{\beta}^* = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}(y_{1,it} - u_{1,it})$ and $\mathbf{V}^* = s^2\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'\right)^{-1}$, where $s^2 = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{1,it} - u_{1,it} - \mathbf{x}_{it}\widehat{\boldsymbol{\beta}}\right)^2$. Then, we can use either an independence or a random walk Metropolis-Hastings algorithm. The benefit is that we avoid the costly inversion of the GLS variance-covariance matrix in each MCMC iteration.

The posterior conditional distribution of $\mathbf{u}_{it}$ is given by

$$p\left(\mathbf{u}_{it}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Xi}\right) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{u}_{it} - \boldsymbol{\mu}_{it}^*\right)'\mathbf{V}^{-1}\left(\mathbf{u}_{it} - \boldsymbol{\mu}_{it}^*\right)\right\} \times \mathbb{1}\left(\mathbf{u}_{it} \geq \mathbf{0}\right), \tag{A.6}$$

or $\mathbf{u}_{it}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\mu}, \boldsymbol{\Xi} \sim \mathbb{N}_d^+(\boldsymbol{\mu}_{it}^*, \mathbf{V})$. Random draws can be obtained, say, in the bivariate case by using the conditional distributions as follows:

$$\begin{aligned} u_{1,it}|u_{2,it}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\Xi} &\sim \mathbb{N}_1^+\left(\widehat{u}_{1,it}, 1/V_{11}\right) \\ u_{2,it}|u_{1,it}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\Xi} &\sim \mathbb{N}_1^+\left(\widehat{u}_{2,it}, 1/V_{22}\right), \end{aligned} \tag{A.7}$$

where $\widehat{u}_{1,it} = \mu_{1,it}^* + \frac{V_{12}}{V_{11}}\left(\mu_{2,it}^* - u_{2,it}\right)$, $\widehat{u}_{2,it} = \mu_{2,it}^* + \frac{V_{12}}{V_{22}}\left(\mu_{1,it}^* - u_{1,it}\right)$, $\mathbf{V}^{-1} = [V_{ij};\ i,j = 1,2]$ and $\boldsymbol{\mu}_{it}^* = \left[\mu_{1,it}^*, \mu_{2,it}^*\right]'$.

Since these distributions are univariate, we can use standard procedures for generating normal random variables truncated from below at zero. Here, we use the acceptance sampling based on an exponential distribution. For the truncated normal distribution $u \sim \mathbb{N}^+(M, v)$, the parameter of the exponential distribution is $\lambda = \frac{M + \sqrt{M^2 + 4v}}{2}$. The draw $u \sim \mathbb{E}\text{xp}(\lambda)$ is accepted with the probability

$$\exp\left\{\lambda u - 1 - \frac{1}{2v}(u - M)^2 + \left(\lambda^{-1} - M\right)^2\right\}, \tag{A.8}$$

which corresponds to the unique solution of the saddle-point problem:

$$\min_{\lambda}\max_{u} : \lambda^{-1}\exp\left\{\lambda u - \frac{1}{2v}(u - M)^2\right\} \times \mathbb{1}\left(u \geq 0\right). \tag{A.9}$$

The analogous problem in the multivariate case $\boldsymbol{u} \sim \mathbb{N}_d^+(\boldsymbol{M}, \boldsymbol{V})$ has the solution $\boldsymbol{\lambda} + \boldsymbol{V}^{-1}\boldsymbol{M} - \boldsymbol{V}^{-1}\boldsymbol{\Lambda}^{-1} = \mathbf{0}_d$, where $\boldsymbol{\Lambda} = \operatorname{diag}\{\boldsymbol{\lambda}\} = \operatorname{diag}\{\lambda_1, ..., \lambda_m\}$ is the matrix of the parameters of the exponential distributions. Assuming $\lambda_1 = \cdots = \lambda_m = \alpha$, the unique solution is given by $\alpha = \left(d^{-1}\sum_{i=1}^{d} u_i^*\right)^{-1}$, where $\boldsymbol{u}^*$ solves the following equations:

$$\mathbf{1}_d = d\left(\mathbf{1}_d'\boldsymbol{u}^*\right)\boldsymbol{V}^{-1}\left(\boldsymbol{u}^* - \boldsymbol{M}\right). \tag{A.10}$$

Next, the posterior conditional distribution of $\boldsymbol{\Sigma}$ is given by

$$p\left(\boldsymbol{\Sigma}|\boldsymbol{\beta}, \boldsymbol{\Omega}, \mathbf{u}, \boldsymbol{\Xi}\right) \propto |\boldsymbol{\Sigma}|^{-nT/2} \exp\left\{-\frac{1}{2} tr\left[\mathbf{A}^*\boldsymbol{\Sigma}^{-1}\right]\right\}, \tag{A.11}$$

where $\mathbf{A}^* = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{r}_{it} - \mathbf{u}_{it}\right)\left(\mathbf{r}_{it} - \mathbf{u}_{it}\right)'$, i.e., a Wishart distribution.

Further, the posterior conditional distribution of $\boldsymbol{\Omega}^{-1}$ is

$$p\left(\boldsymbol{\Omega}^{-1}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\Xi}\right) \propto |\boldsymbol{\Omega}|^{-nT/2} \exp\left\{-\frac{1}{2} tr\left[\mathbf{A}^{**}\boldsymbol{\Omega}^{-1}\right]\right\} \Phi_\rho\left(\boldsymbol{\Omega}_{diag}^{-1}\boldsymbol{\mu}\right)^{-nT}, \tag{A.12}$$

where $\mathbf{A}^{**} = \sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{u}_{it} - \boldsymbol{\mu}\right)\left(\mathbf{u}_i - \boldsymbol{\mu}\right)'$, which would have been a Wishart distribution if it were not for the last term.

Finally, we have

$$p\left(\boldsymbol{\mu}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \mathbf{u}, \boldsymbol{\Xi}\right) \propto \exp\left\{-\frac{1}{2} tr\left[\boldsymbol{\Omega}^{-1}\right]\mathbf{A}^{**}\right\} \Phi_\rho\left(\boldsymbol{\Omega}_{diag}^{-1}\boldsymbol{\mu}\right)^{-nT}, \tag{A.13}$$

from where it follows that $\boldsymbol{\mu}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\Xi} \sim \mathbb{N}_d\left((nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{u}_{it},\ (nT)^{-1}\boldsymbol{\Omega}\right)$, apart from the last term in the above expression.

We note that, if the location parameter of the truncated normal distribution of $\mathbf{u}_{it}$ is not constant but varying with some contextual variables, i.e., if $\mu_{it} = \begin{bmatrix}\mu_{1,it}\\ \mu_{2,it}\end{bmatrix} = \begin{bmatrix}\mathbf{z}_{1,it}'\boldsymbol{\gamma}_1\\ \mathbf{z}_{2,it}'\boldsymbol{\gamma}_2\end{bmatrix} = \begin{bmatrix}\mathbf{z}_{1,it} & \mathbf{0}_{l_1}\\ \mathbf{0}_{l_2} & \mathbf{z}_{2,it}\end{bmatrix}' \begin{bmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\gamma}_2\end{bmatrix} \equiv \mathbf{Z}_{it}'\boldsymbol{\gamma}$, where $\mathbf{z}_{1,it}$ and $\mathbf{z}_{2,it}$ is an $l_1 \times 1$ and $l_2 \times 1$ vector of covariates, and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ is an $l_1 \times 1$ and $l_2 \times 1$ vector of the corresponding parameters, respectively, then

$$\begin{aligned} p\left(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \mathbf{u}, \boldsymbol{\Xi}\right) \propto\ & \exp\left\{-\frac{1}{2} tr\left[\boldsymbol{\Omega}^{-1}\right]\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{u}_{it} - \mathbf{Z}_{it}'\boldsymbol{\gamma}\right)\left(\mathbf{u}_{it} - \mathbf{Z}_{it}'\boldsymbol{\gamma}\right)'\right\} \times \\ & \prod_{i=1}^{n}\prod_{t=1}^{T}\Phi_\rho\left(\boldsymbol{\Omega}_{diag}^{-1}\mathbf{Z}_{it}'\boldsymbol{\gamma}\right)^{-1}. \end{aligned} \tag{A.14}$$

From the first half of the above expression it follows that $\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \mathbf{u}, \boldsymbol{\Xi} \sim \mathbb{N}_{l_1+l_2}\left(\widehat{\boldsymbol{\gamma}}, \mathbf{V}_\gamma\right)$, where $\widehat{\boldsymbol{\gamma}} = \left(\mathbf{Z}'\left(\mathbf{I}_{nT} \otimes \boldsymbol{\Omega}^{-1}\right)\mathbf{Z}\right)^{-1}\mathbf{Z}'\left(\mathbf{I}_{nT} \otimes \boldsymbol{\Omega}^{-1}\right)\mathbf{U}$ and $\mathbf{V}_\gamma = \left(\mathbf{Z}'\left(\mathbf{I}_{nT} \otimes \boldsymbol{\Omega}^{-1}\right)\mathbf{Z}\right)^{-1}$ with $\mathbf{U} = [u_{1,11}, \ldots, u_{1,nT}, u_{2,11}, \ldots, u_{2,nT}]'$ and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_{1,11}' & \mathbf{0} \\ \vdots & \vdots \\ \mathbf{z}_{1,nT}' & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{2,11}' \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{z}_{2,nT}' \end{bmatrix}$$

being of the $2nT \times 1$ and $2nT \times (l_1 + l_2)$ dimensions, respectively. Apart from the nonlinear product term appearing in the second half of the expression in (A.14), the posterior conditional corresponds to the SUR model. Hence, for samples of the large size, $\widehat{\boldsymbol{\gamma}}$ can be computed as $\widehat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Omega}^{-1}\mathbf{Z}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Omega}^{-1}\mathbf{U}_{it}$ along with $\mathbf{V}_{\gamma} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Omega}^{-1}\mathbf{Z}_{it}'\right)^{-1}$.

B Model II

Since, in our empirical application, we set $\mathbf{z}_{1,it} = \mathbf{z}_{2,it} = \left[1, t, \frac{1}{2}t^2\right]'$, here we focus on the special case when $\mathbf{z}_{1,it} = \mathbf{z}_{2,it} \equiv \mathbf{z}_{it}$, which is said to be of dimension l. When estimating the hierarchical model, the only component which changes in the MCMC scheme is the way we sample the latent variables and their corresponding parameters. After some algebra in the kernel posterior distribution, we derive the following posterior conditional distributions:

$$\mathbf{u}_{it}|\boldsymbol{\beta},\boldsymbol{\gamma},\psi_1,\boldsymbol{\Sigma},\omega_1^2,\omega_2^2,\boldsymbol{\Xi} \;\sim\; \mathbb{N}_2^{+}\left(\widehat{\mathbf{u}}_{it},\mathbf{V}_u\right)\times\Phi\left(\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_1+\psi_1 u_{2,it}\right)/\omega_1\right)^{-1}, \tag{B.1}$$

where

$$\begin{aligned}
\widehat{\mathbf{u}}_{it} &= \left(\omega_1^2\omega_2^2\boldsymbol{\Sigma}^{-1}+\left(\omega_1^2+\omega_2^2\left(1+\psi_1^2\right)\right)\mathbf{I}_2\right)^{-1}\left(\omega_1^2\omega_2^2\boldsymbol{\Sigma}^{-1}\mathbf{r}_{it}+[\boldsymbol{\iota}_2\otimes\mathbf{z}_{it}]'\boldsymbol{\varphi}\right)\\
\boldsymbol{\varphi} &= \begin{bmatrix}\boldsymbol{\gamma}_1/\omega_2^2\\ \omega_1^2\left(\boldsymbol{\gamma}_2-\psi_1\omega_2^2\boldsymbol{\gamma}_1\right)\end{bmatrix}\\
\mathbf{V}_u &= \omega_1^2\omega_2^2\left(\boldsymbol{\Sigma}^{-1}+\left(\omega_1^2+\omega_2^2\left(1+\psi_1^2\right)\right)\mathbf{I}_2\right)^{-1},
\end{aligned}$$

where $\boldsymbol{\iota}_2$ is a 2×1 vector of ones.

Since $u_{2,it}$ appears in a nonlinear way, we can use the following alternative posterior conditional distributions:

$$u_{1,it}|u_{2,it},\boldsymbol{\beta},\boldsymbol{\gamma},\psi_1,\boldsymbol{\Sigma},\omega_1^2,\omega_2^2,\boldsymbol{\Xi} \;\sim\; \mathbb{N}^{+}\left(\mu_{1,it},\;\frac{\omega_1^2}{1+\sigma_{11}\omega_1^2}\right) \tag{B.2a}$$

$$u_{2,it}|u_{1,it},\boldsymbol{\beta},\boldsymbol{\gamma},\psi_1,\boldsymbol{\Sigma},\omega_1^2,\omega_2^2,\boldsymbol{\Xi} \;\sim\; \mathbb{N}^{+}\left(\mu_{2,it},\;\frac{\omega_2^2}{1+\sigma_{22}\omega_2^2}\right)\times\Phi\left(\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_1+\psi_1 u_{2,it}\right)/\omega_1\right)^{-1}, \tag{B.2b}$$

where

$$\mu_{1,it} = \frac{\left(\psi_1-\omega_1^2\sigma_{12}\right)u_{2,it}+\mathbf{z}_{it}'\boldsymbol{\gamma}_1+\omega_1^2\left(\boldsymbol{\Sigma}^{-1}\mathbf{r}_{it}\right)_1}{1+\sigma_{11}\omega_1^2}, \qquad \mu_{2,it} = \frac{-\omega_2^2\sigma_{12}u_{1,it}+\mathbf{z}_{it}'\boldsymbol{\gamma}_2+\omega_2^2\left(\boldsymbol{\Sigma}^{-1}\mathbf{r}_{it}\right)_2}{1+\sigma_{22}\omega_2^2}$$

and $\left(\boldsymbol{\Sigma}^{-1}\mathbf{r}_{it}\right)_j$ denotes the jth row of $\left(\boldsymbol{\Sigma}^{-1}\mathbf{r}_{it}\right)$ for $j=1,2$. Although the first conditional distribution above is a truncated normal, the second one is however not due to the presence of its last term, which comes from the normalizing constant of the prior conditional for $u_{1,it}|u_{2,it},\mathbf{z}_{it}$. To draw samples from this conditional distribution, we first write it in the form of $x\sim\mathbb{N}^{+}\left(\mu,v^2\right)\times\Phi\left(a+bx\right)^{-1}$ using obvious notation. Suppose $w = a+bx$. It is then easy to show that the derivatives of the log density are $f'(x) = \frac{1}{2}(bv)^{-1}(w-a-b\mu)+bv\Lambda(w)$, where $\Lambda(w) = \phi(w)/\Phi(w)$ and $-f''(x) = \frac{1}{2}+(bv)^2\Lambda(w)\left[1+\Lambda(w)\right] > 0$ for all $w\in\mathbb{R}$, from where it follows that the distribution is log-concave. The mode satisfies the following nonlinear equation: $w^*+2(bv)^2\Lambda(w^*)-(a+b\mu)=0$, from where we get $x^* = b^{-1}\left(w^*-a\right)$. Thus, we can use acceptance sampling when the source distribution is $x\sim\mathbb{N}^{+}\left(x^*,-f''\left(x^*\right)^{-1}\right)$.

When the sample size is large, we have to resort to certain simplifications to speed up the procedure. First, to avoid solving the nonlinear equation, we set $x\sim\mathbb{N}^{+}\left(x^*,-f''\left(x^*\right)^{-1}\right)$ which

is one fixed-point iteration away from the initial condition $w^{(0)} = a + b\mu$. Second, in extreme cases when it takes more than 100 rejections to obtain a draw, we resort to a Metropolis-Hastings by drawing $x \sim \mathbb{N}^+\left(a^{-1}\left(w^* - b\right), -f''\left(a^{-1}\left(w^* - b\right)\right)^{-1}\right)$ and accepting the draw with the probability

$$\min\left\{1,\ \exp\left\{-\tfrac{1}{2v^2}\left[(x-\mu)^2 + \left(x^{(o)} - \mu\right)^2\right] + \tfrac{1}{2v^2}\left[(x-x^*)^2 + \left(x^{(o)} - x^*\right)^2\right]\right\}\Phi\left(a+bx\right)^{-1}\Phi\left(a+bx^{(o)}\right)\right\}. \tag{B.3}$$

For other parameters, the posterior conditional distributions are as follows:

$$\begin{bmatrix}\boldsymbol{\gamma}_1 \\ \psi_1\end{bmatrix}\Big|\mathbf{u}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma}_2, \boldsymbol{\Sigma}, \omega_1^2, \omega_2^2, \boldsymbol{\Xi} \ \sim\ \mathbb{N}_{l+1}\left(\left(\mathbf{Z}_u'\mathbf{Z}_u\right)^{-1}\mathbf{Z}_u'\mathbf{u}_1,\ \omega_1^2\left(\mathbf{Z}_u'\mathbf{Z}_u\right)^{-1}\right) \times \prod_{i=1}^{n}\prod_{t=1}^{T}\Phi\left(\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_1 + \psi_1 u_{2,it}\right)/\omega_1\right)^{-1} \tag{B.4a}$$

$$\boldsymbol{\gamma}_2|\mathbf{u}_{it}, \boldsymbol{\beta}, \psi_1, \boldsymbol{\Sigma}, \omega_1^2, \omega_2^2, \boldsymbol{\Xi} \ \sim\ \mathbb{N}_l\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{u}_2,\ \omega_2^2(\mathbf{Z}'\mathbf{Z})^{-1}\right) \times \prod_{i=1}^{n}\prod_{t=1}^{T}\Phi\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_2/\omega_2\right)^{-1}, \tag{B.4b}$$

where $\mathbf{Z} = [\mathbf{z}_{11}, \ldots, \mathbf{z}_{nT}]'$, $\mathbf{u}_1 = [u_{1,11}, \ldots, u_{1,nT}]'$, $\mathbf{u}_2 = [u_{2,11}, \ldots, u_{2,nT}]'$ and $\mathbf{Z}_u = \begin{bmatrix}\mathbf{Z} & \mathbf{u}_2\end{bmatrix}$.

C Model III

Similar to Model II, here we focus on the special case when $\mathbf{z}_{1,it} = \mathbf{z}_{2,it} \equiv \mathbf{z}_{it}$, which is said to be of dimension l. If approximations to the marginal distributions are available, then the joint distribution of $u_{1,it}$ and $u_{2,it}$ is

$$\widetilde{p}\left(u_{1,it}, u_{2,it}|\mathbf{z}_{it}\right) = \left[p\left(u_{1,it}|u_{2,it}, \mathbf{z}_{it}\right)p\left(u_{2,it}|u_{1,it}, \mathbf{z}_{it}\right)\widetilde{p}\left(u_{1,it}|\mathbf{z}_{it}\right)\widetilde{p}\left(u_{2,it}|\mathbf{z}_{it}\right)\right]^{1/2}, \tag{C.1}$$

where $\widetilde{p}\left(u_{1,it}|\mathbf{z}_{it}\right)$ and $\widetilde{p}\left(u_{2,it}|\mathbf{z}_{it}\right)$ denote certain approximations to the marginal distributions. Specifically, we use the following approximations:

$$u_{1,it}|\mathbf{z}_{it} \ \sim\ \mathbb{N}^+\left(\widehat{\mu}_{1,it},\ \widehat{\omega}_1^2\right) \tag{C.2a}$$

$$u_{2,it}|\mathbf{z}_{it} \ \sim\ \mathbb{N}^+\left(\widehat{\mu}_{2,it},\ \widehat{\omega}_2^2\right), \tag{C.2b}$$

where the location and scale parameters are to be determined. To build such approximations, we use MCMC to obtain a large sample of dependent draws $\left\{\mathbf{u}_{it}^{(s)} = \left[u_{1,it}^{(s)}, u_{2,it}^{(s)}\right]',\ s = 1, ..., S\right\}$ from the specification of conditional distributions. Therefore, the posterior distribution of the latent variables is

$$\begin{aligned}
p\left(u_{1,it}, u_{2,it}|\boldsymbol{\Xi}, \boldsymbol{\theta}\right) &= \exp\left\{-\frac{1}{2}\left(\mathbf{u}_{it} - \mathbf{r}_{it}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{u}_{it} - \mathbf{r}_{it}\right)\right\} \times \\
&\quad \left[p\left(u_{1,it}|u_{2,it}, \mathbf{z}_{it}\right)p\left(u_{2,it}|u_{1,it}, \mathbf{z}_{it}\right)\widetilde{p}\left(u_{1,it}|\mathbf{z}_{it}\right)\widetilde{p}\left(u_{2,it}|\mathbf{z}_{it}\right)\right]^{1/2} \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{u}_{it} - \mathbf{r}_{it}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{u}_{it} - \mathbf{r}_{it}\right) - \frac{1}{4\omega_1^2}\left(u_{1,it} - \mathbf{z}_{it}'\boldsymbol{\gamma}_1 - \psi_1 u_{2,it}\right)^2 - \frac{1}{4\omega_2^2}\left(u_{2,it} - \mathbf{z}_{it}'\boldsymbol{\gamma}_2 - \psi_2 u_{2,it}\right)^2\right\} \times \\
&\quad \Phi^{-1/2}\left(\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_1 + \psi_1 u_{2,it}\right)/\omega_1\right)\Phi^{-1/2}\left(\left(\mathbf{z}_{it}'\boldsymbol{\gamma}_2 + \psi_2 u_{1,it}\right)/\omega_2\right) \times \\
&\quad \exp\left\{-\frac{1}{4\widehat{\omega}_1^2}\left(u_{1,it} - \widehat{\mu}_{1,it}\right)^2 - \frac{1}{4\widehat{\omega}_2^2}\left(u_{2,it} - \widehat{\mu}_{2,it}\right)^2\right\}.
\end{aligned} \tag{C.3}$$

Now, we need to determine the location and scale parameters in the approximations of the marginal distributions. Based on the univariate approximately random samples $\left\{u_{1,it}^{(s)},\ s=1,...,S\right\}$ and $\left\{u_{2,it}^{(s)},\ s=1,...,S\right\}$ obtained by MCMC from the set of conditional distributions, we can employ the ML method to determine parameters $\widehat{\zeta}_1$, $\widehat{\zeta}_2$, $\widehat{\omega}_1^2$ and $\widehat{\omega}_2^2$ under the assumption that $\widehat{\mu}_{1,it}=\mathbf{z}_{it}'\widehat{\zeta}_1$ and $\widehat{\mu}_{2,it}=\mathbf{z}_{it}'\widehat{\zeta}_2$. The ML estimates can be obtained by maximizing the following log-likelihood:

$$\mathscr{L}_j=-\frac{nT}{2}\ln\widehat{\omega}_j^2-\frac{1}{2\widehat{\omega}_j^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{u}_{j,it}-\mathbf{z}_{it}'\widehat{\zeta}_j\right)^2-\sum_{i=1}^{n}\sum_{t=1}^{T}\ln\Phi\left(\mathbf{z}_{it}'\widehat{\zeta}_j/\widehat{\omega}_j\right)\quad\forall\ j=1,2. \tag{C.4}$$

Then, from the expression

$$p\left(u_{1,it},u_{2,it}|\boldsymbol{\Xi},\boldsymbol{\theta}\right)=\exp\left\{-\frac{1}{2}\left(\mathbf{u}_{it}-\mathbf{r}_{it}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{u}_{it}-\mathbf{r}_{it}\right)\right\}p\left(u_{1,it},u_{2,it}|\mathbf{z}_{it}\right) \tag{C.5}$$

we can obtain, through a small-scale MCMC, a draw $\{\widetilde{\mathbf{u}}_{it},\ i=1,...,n;\ t=1,\ldots,T\}$ from $p\left(u_{1,it},u_{2,it}|\mathbf{z}_{it}\right)]$. Given the existing draw $\left\{\mathbf{u}_{it}^{(o)},\ i=1,...,n;\ t=1,\ldots,T\right\}$, the new draw will be accepted with the probability

$$\min\left\{1,\ \exp\left\{-\frac{1}{2}tr\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{\mathbf{u}}_{it}-\mathbf{r}_{it}\right)\left(\widetilde{\mathbf{u}}_{it}-\mathbf{r}_{it}\right)'-\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{u}_{it}^{(o)}-\mathbf{r}_{it}\right)\left(\mathbf{u}_{it}^{(o)}-\mathbf{r}_{it}\right)'\right]\right\}\right\}. \tag{C.6}$$

This procedure updates the latent variables as a group for all observations so if the acceptance probability is not exceedingly low or exceedingly high, it is expected to perform quite well.

References

Berger, A. (1995). The profit-structure relationship in banking — Tests of market-power and efficient-structure hypotheses. *Journal of Money, Credit, and Banking*, *27*, 404–431.

Berger, A. & Hannan, T. (1998). The efficiency cost of market power in the banking industry: A test of the "quiet life" and related hypotheses. *Review of Economics and Statistics*, *80*, 454–465.

Berger, A., Klapper, L. F., & Turk-Ariss, R. (2009). Bank competition and financial stability. *Journal of Financial Services Research*, *35*, 99–118.

Berger, A. N. & Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking & Finance*, *21*(7), 895–947.

Berger, A. N. & Mester, L. J. (2003). Explaining the dramatic changes in performance of US banks: Technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation*, *12*(1), 57–95.

Bolt, W. & Humphrey, D. (2015). A frontier measure of U.S. banking competition. *European Journal of Operational Research*, *246*, 450–461.

Boyd, J. H. & De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. *Journal of Finance*, *60*, 1329–1343.

Brissimis, S. N. & Delis, M. D. (2011). Bank-level estimates of market power. *European Journal of Operational Research*, *212*, 508–517.

Carbo, S., Humphrey, D., Maudos, J., & Molyneux, P. (2009). Cross-country comparisons of competition and pricing power in European banking. *Journal of International Money and Finance*, *28*, 115–134.

Casu, B. & Girardone, C. (2006). Does competition lead to efficiency? The case of EU commercial banks. Working Paper, Cass Business School, City University.

Chirinko, R. S. & Fazzari, S. M. (2000). Market power and inflation. *Review of Economics and Statistics*, *82*, 509–518.

Claessens, S. & Laeven, L. (2005). Financial dependence, banking sector competition, and economic growth. *Journal of the European Economic Association*, *3*, 179–207.

Das, A. & Kumbhakar, S. C. (2016). Markup and efficiency of Indian banks: An input distance function approach. *Empirical Economics*. forthcoming.

de Guevara, F. & Maudos, J. (2007). On the relationship between competition and efficiency in the EU banking sectors. *Manchester School*, *75*, 275–296.

Delis, M., Iosifidi, M., & Tsionas, E. G. (2014). On the estimation of marginal cost. *Operations Research*, *62*, 543–556.

Delis, M. & Tsionas, E. G. (2009). The joint estimation of bank-level market power and efficiency. *Journal of Banking and Finance*, *33*, 1842–1850.

Demsetz, H. (1973). Industry structure, market rivalry and public policy. *Journal of Law and Economics*, *16*, 1–19.

Demsetz, R., Saidenberg, M., & Strahan, P. (1996). Banks with something to lose: The disciplinary role of franchise value. *Economic Policy Review*, *2*.

DeYoung, R. & Rice, T. (2004). How do banks make money? A variety of business strategies. *Economic Perspectives - Federal Reserve Bank of Chicago*, *28(4)*.

DiCiccio, T. J., Kass, R. E., Raftery, A., & Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, *92*, 903–915.

Dong, Y., Firth, M., Hou, W., & Yang, W. (2016). Evaluating the performance of Chinese commercial banks: A comparative analysis of different types of banks. *European Journal of Operational Research*, *252*, 280–295.

Gan, J. (2004). Banking market structure and financial stability: Evidence from the Texas real estate crisis in the 1980s. *Journal of Financial Economics*, *73*, 567–601.

Goldberg, L. G. & Rai, A. (1996). The structure – performance relationship for European banking. *Journal of Banking and Finance*, *20*, 745–771.

Hannan, T. H. & Prager, R. A. (2004). The competitive implications of multimarket bank branching. *Journal of Banking and Finance*, *28*, 1889–1914.

Hannan, T. H. & Prager, R. A. (2009). The profitability of small single-market banks in an era of multimarket banking. *Journal of Banking and Finance*, *33*, 263–271.

Hicks, J. R. (1935). Annual survey of economic theory: The theory of monopoly. *Econometrica*, *3*, 1–20.

Hughes, J. P. & Mester, L. J. (2013). Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation*, *22*(4), 559–585.

Jayaratne, J. & Strahan, P. E. (1998). Entry restrictions, industry evolution, and dynamic efficiency: Evidence from commercial banking. *Journal of Law and Economics*, *41*, 239–273.

Koetter, M., Kolari, J. W., & Spierdijk, L. (2012). Enjoying the quiet life under deregulation? Evidence from adjusted Lerner indices for U.S. banks. *Review of Economics and Statistics*, *94*, 462–480.

Koetter, M. & Poghosyan, T. (2009). The identification of technology regimes in banking: Implications for the market power – fragility nexus. *Journal of Banking and Finance*, *33*, 1413–1422.

Koetter, M. & Vins, O. (2008). The quiet life hypothesis in banking – Evidence from German savings banks. Working Paper Series: Finance and Accounting, Johann Wolfgang Goethe – Universität Frankfurt am Main.

Laeven, L. & Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, *93*, 259–275.

Malikov, E., Kumbhakar, S. C., & Tsionas, M. G. (2016). A cost system approach to the stochastic directional

technology distance function with undesirable outputs: The case of U.S. banks in 2001–2010. *Journal of Applied Econometrics*, *31*, 1407–1429.

Malikov, E., Restrepo-Tobón, D., & Kumbhakar, S. C. (2015). Estimation of banking technology under credit uncertainty. *Empirical Economics*, *49*(1), 185–211.

Maudos, J. & Fernández de Guevara, J. (2007). The cost of market power in the European banking sectors: Social welfare cost vs. cost inefficiency. *Journal of Banking and Finance*, *31*, 2103–2125.

Molyneux, P., Lloyd-Williams, D. M., & Thornton, J. (1994). Competitive conditions in European banking. *Journal of Banking and Finance*, *18*, 445–459.

Pitt, M. M. & Lee, L.-F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics*, *9*, 43–64.

Punt, L. W. & van Rooij, M. C. J. (1999). The profit – structure relationship, efficiency and mergers in the European banking industry: An empirical assessment. Research Memorandum WO&E no. 604, De Nederlandsche Bank.

Restrepo-Tobón, D. & Kumbhakar, S. C. (2014). Enjoying the quiet life under deregulation? Not quite. *Journal of Applied Econometrics*, *29*, 333–343.

Schaeck, K. & Cihák, M. (2008). How does competition affect efficiency and soundness in banking? New empirical evidence. Working Paper No. 932, European Central Bank.

Sealey, C. W. & Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance*, *32*, 1251–1266.

Stiroh, K. J. & Strahan, P. E. (2003). Competitive dynamics of competition: Evidence from U.S. banking. *Journal of Money, Credit, and Banking*, *35*, 801–828.

Turk Ariss, R. (2010). On the implications of market power in banking: Evidence from developing countries. *Journal of Banking and Finance*, *34*, 765–775.

Weill, L. (2004). On the relationship between competition and efficiency in the EU banking sectors. *Kredit und Kapital*, *37*, 329–352.

Wheelock, D. C. & Wilson, P. W. (2012). Do large banks have lower costs? New estimates of returns to scale for US banks. *Journal of Money, Credit and Banking*, *44*(1), 171–199.