An Anatomy of Industry Merger Waves[*]

Daniele Bianchi[†]   Carlo Chiarella[‡]

Abstract

We propose a novel panel Bayesian Markov regime-switching Poisson regression model with time-varying transition probabilities to test existing theories on the driving forces of wave-like patterns in same-industry mergers and acquisitions (M&A). We show that the dynamics and persistence of merger waves change substantially in the cross-section of deal flow. This suggests that any inference on existing economically justified competing explanations of merger waves at the aggregate market level could be misleading, as the observed cross-industry heterogeneity in waves is shown to be the consequence of different responses to common or distinct drivers of merger activity.

Keywords: Mergers and Acquisitions, Poisson Regression, Markov Regime-Switching, Time-Varying Probabilities, Markov Chain Monte Carlo.

JEL codes: C11, C22, C24, C58

1 Introduction

Periodic waves of mergers and acquisitions (M&A) have been an integral part of capitalist development since its inception, representing an important mechanism for the reorganization and restructuring of a market economy across industries. Economic theory suggests that mergers can occur either in response to economic shocks that trigger restructuring and consolidation, or be driven by managerial market timing. Nonetheless, to date, there is no clear consensus as to why merger deals have occurred in cyclical patterns in which periods of intense activity have been followed by intervening periods of fewer mergers.

In this paper, we define a merger wave as a substantial, yet temporary, increase in the intensity of M&A activity, and develop econometric tools to investigate whether the dominant view that merger waves have different dynamics across industries is backed up by empirical evidence (see, for example, Mitchell and Mulherin, 1996). Then, we also ask whether wave-like patterns of merger activity at the industry level may derive from heterogeneous exposure to competing theory-based explanatory variables, such as sector-specific economic shocks, managers' market timing, or aggregate macro-financial conditions.

In methodological terms, we propose a novel Markov regime-switching Poisson regression approach with time-varying transition probabilities that allows for the explicit investigation of the driving factors behind abnormal M&A activity. The main underlying assumption of the model is that the decision to embark in M&A depends on a set of parameters that represent the reaction of agents to realizations of a latent industry-specific discrete state S_t, which identifies mounting evidence of deals. The intuition is simple: from the point of view of the corporate decision maker, waves in M&A at time t are not determined by her own decision. However, observing abnormally high merger activity increases the aggregate propensity to engage in a deal. As a result, the question of understanding the driving factors in the dynamics of S_t coincides with investigating why managers value mergers more in some states than others. This modeling framework is fairly general and implies that the intensity of merger activity can differ across industries, that it is multi-factorial in nature, and that it is unlikely to be entirely captured by any single, observable, industry-specific characteristic or aggregate factor. We model merger waves as a first-order Markov state with time-varying transition probabilities that accommodates the time-varying effect of multiple industry-specific and aggregate

macro-financial factors. The cross-sectional heterogeneity and correlations across different industries are captured by industry-specific random-effects that are sampled from a common distribution.

A few measurement tools for merger waves have been proposed in the literature (see, for example, Shugart and Tollison, 1984, Town, 1992, Golbe and White, 1993, Barkoulas et al., 2001, Resende, 2008, and Choi and Jeon, 2011). The common feature of these approaches is that determinants of merger waves are investigated in a conditionally Gaussian framework and separately after a latent state S_t is identified. As a result, any causality statement is thus to be read as contingent on having full confidence on the normal assumption for the distribution of M&A activity and the consequential identification of waves. Unless additional assumptions are introduced, such procedures can potentially suffer from endogeneity and misspecification issues that can distort standard hypothesis testing. We extend the existing literature by proposing a model that explicitly and jointly acknowledges the time-varying, uncertain, and non-Gaussian nature of M&A activity.

As far as the model estimation is concerned, we follow Chib and Winkelmann (2001), Frühwirth-Schnatter and Wagner (2006), Frühwirth-Schnatter and Frühwirth (2007) and Kaufmann (2015) and implement an approximate, yet accurate, Markov Chain Monte Carlo (MCMC) estimation scheme for the unknown parameters, the hidden discrete states identifying merger waves, and the time-varying transition probabilities. By using an MCMC approach, we are able to make inferences on parameters and latent states in a single step, developing a finite-sample testing framework that is based on the posterior distributions of virtually any function of the model outputs. This allows us to test hypotheses on the nature, dynamics, and structure of industry merger waves in a unified setting, which is something that earlier literature did not do.

Empirically, we focus on a sample of M&A bids announced by US private and public acquirers for US public and private targets in the period from 1983 to 2016. M&A deals are aggregated at the industry level for each quarter, assigning each deal to one of 12 industries based on the four-digit SIC code of the bidder at the time of the announcement. We test a set of common competing explanatory variables of merger waves; first, we consider industry-specific stock valuation variables related to the behavioural hypothesis, which suggests that M&A activity is driven by firm-specific and market-wide valuations, or managerial market timing (see, for example, Shleifer and Vishny, 2003, Rhodes-Kropf and Viswanathan, 2004, Rhodes-Kropf et al., 2005 and Maksimovic et al., 2013). A second set of explanatory variables is related to the neoclassical hypothesis, which suggests M&A activity primarily

3

occurs in response to technological, regulatory and/or economic shocks that trigger restructuring and consolidation at the industry level (see, for example, Coase, 1937, Gort, 1969, Shleifer and Vishny, 1992, Harford, 1999, Andrade et al., 2001, Holmstrom and Kaplan, 2001, Maksimovic and Phillips, 2001, Jovanovic and Rousseau, 2002, and Harford, 2005). In addition to these mainstream theories, we also investigate the explanatory power of aggregate macro-financial conditions (see, for example, Melicher et al., 1983, Shugart and Tollison, 1984, Becketti, 1986, Town, 1992, Golbe and White, 1993, Mulherin and Boone, 2000, and Choi and Jeon, 2011). Although with mixed results, these theories have been found to explain M&A activity over the last century, and thus are clearly relevant to a comprehensive understanding of merger waves.

We report two main novel results. First, we show that the dynamics and persistence of merger waves change substantially across industries over the last thirty years. This suggests that wave-like patterns are essentially industry specific. As a matter of fact, we show that the posterior estimates of the cross-sectional correlations of industry-specific random-effects are rather low, albeit significant. Second, we show that such heterogeneity can be explained by differences in the transition dynamics that drive the states of merger waves. For instance, the uncertainty around valuations represents a significant explanatory variable that negatively affects M&A activity for industries such as Manufacturing, Telecommunications, and Healthcare, while it does not appear to be relevant for, say, the Money industry. Similarly, an index of economic activity, as constructed in Harford (2005), turns out to significantly and negatively affect M&A only in the financial sector, while it barely appears in other industries. Interestingly, merging activity in the Money industry is tightly linked with capital liquidity, consistent with the idea that, in the financial industry, M&A deals are mostly pro-cyclical and tend to occur for restructuring purposes. As a whole, the results show that deteriorating aggregate economic conditions and the uncertainty related to market valuations are expected to be associated with the intensity of M&A deals.

2   Modeling Merger Waves

The number of deals y_{it} for each industry $i = 1, \ldots, n$ and time $t = 1, \ldots, T$ is modeled as a Poisson process in which the intensity rate is a direct function of a set of industry-specific variables $\mathbf{w}_{it}$, macroeconomic information $\mathbf{x}_t$, and a parameter μ_i that represents the unobserved cross-sectional

heterogeneity, i.e, a group-specific random effect,

$$y_{it}|\lambda_{it} \sim Poisson\left(\lambda_{it}\right), \qquad \lambda_{it} = \exp\left(\mathbf{z}_{it}'\boldsymbol{\beta}_{S_{it}}^{i} + \mu_i\right), \tag{1}$$

where $\mathbf{z}_{it}' = (\mathbf{x}_t', \mathbf{w}_{it}')$ is a p-dimensional vector of explanatory variables, $\boldsymbol{\beta}_k^{i,\prime} = \left(\boldsymbol{\beta}_{1k}^{i}, \ldots, \boldsymbol{\beta}_{pk}^{i}\right)$ is the vector of the corresponding regression parameters for $S_t = k$, $\mu_i \sim N\left(\boldsymbol{\mu}, D\right)$ is a random effect that is sampled from a common multivariate Normal distribution, and S_{it} is a latent discrete state that identifies a merger wave for the ith industry. That is, the intensity of M&A activity is state-dependent. We assume the latent state follows an independent first-order Markov process with time-varying transition probabilities specified through a multi-Logit function, i.e.,

$$p\left(S_{it} = k|S_{it-1} = l, \mathbf{z}_{it-1}, \boldsymbol{\alpha}_i\right) = \pi_{lk,t}^{i} = \frac{\exp\left(\mathbf{z}_{it-1}'\boldsymbol{\alpha}_{ilk}^{z} + \alpha_{ilk}\right)}{\sum_{j=1}^{K} \exp\left(\mathbf{z}_{it-1}'\boldsymbol{\alpha}_{ilj}^{z} + \alpha_{ilj}\right)}, \tag{2}$$

with $\boldsymbol{\alpha}_i' = \left(\boldsymbol{\alpha}_{i,S_{it-1},S_{it}}^{z,\prime}, \boldsymbol{\alpha}_{i,S_{it-1},S_{it}}'\right)$ that is a $K(p+1)-$dimensional vector of parameters. The hidden nature of S_{it} prevents prevents the deteriministic mapping of the explanatory variables $\mathbf{z}_{it-1}$ into the joint decision-making process of bidders and eventually into abnormally high states of merger activity.

The formulation in (2) allows us to decompose the dynamics of merger waves into two components: a time-varying component $\mathbf{z}_{it-1}'\boldsymbol{\alpha}_{i,S_{it-1},S_{it}}^{z}$ that captures the pure effect of the determinants on the regime-switching probabilities, and a time-invariant component $\alpha_{i,S_{it-1},S_{it}}$ that captures the persistence of the latent state independently from exogenous factors. The reference state is defined as the state in which the number of deals cannot be classified as falling within a merger wave. The remaining states represent the increasing strength of merger activity.

For identification purposes, the parameters governing the transition to the state of no merger wave are restricted to be equal to zero, i.e. $\left(\boldsymbol{\alpha}_{i,S_{it-1},1}^{z,\prime}, \boldsymbol{\alpha}_{S_{i,it-1},1}'\right) = (0', 0')$. For example, for a model with two states, this yields:

$$p\left(S_{it} = 1|S_{it-1} = l, \mathbf{z}_{it-1}, \boldsymbol{\alpha}_i\right) = \pi_{l1,t} = \frac{1}{1 + \exp\left(\mathbf{z}_{it-1}'\boldsymbol{\alpha}_{il2}^{z} + \alpha_{il2}\right)}, \qquad l = 1, 2 \tag{3}$$

such that the vector of parameters $\left(\boldsymbol{\alpha}_{il2}^{z}, \alpha_{il2}\right)$ is the only set of parameters we need to estimate. The specification in (2) is rather general as it involves standard implementations as special cases. For instance, by restricting the effect of the covariate $\mathbf{z}_{it-1}$ to be independent of past states, i.e.

5

$\boldsymbol{\alpha}_{ilk}^z = \boldsymbol{\alpha}_{ik}^z$, we assume that the persistence of the latent state is governed only by the time-invariant component α_{ilk} (see, for example, Filardo, 1994, and Amisano and Fagan, 2013). Further, a standard time-homogeneous Markov regime-switching model can be recovered as a special case by restricting $\boldsymbol{\alpha}_{ilk}^z = 0, \forall l, k$. If, instead, we restrict both $\boldsymbol{\alpha}_{ilk}^z = \boldsymbol{\alpha}_{ik}^z$ and $\alpha_{ilk} = \alpha_{ik}$, we obtain the multi-state analogue of the smooth transition model of Terasvirta and Anderson (1992) where regime probabilities are a direct, monotone function of $\mathbf{z}_{it-1}$, and the transition mechanism is independent of the lagged prevailing state.

One comment is in order. Our model captures over-dispersion in the dynamics of deal flow in an unconditional sense, but not conditionally on the latent state of a merger wave. In this respect, the Poisson regression in (1) implies that the unconditional variance of deal flow is higher than the unconditional mean.

However, conditional on the latent state S_{it}, we have that $Var[y_{it}|S_{it}] = E[y_{it}|S_{it}] = \lambda_{S_{it}}$. A possible solution would be to estimate a mixture of negative binomial distributions, which has one parameter more than the Poisson regression and adjusts the variance independently from the mean. We believe that would be an interesting extension to our modeling framework and represent a promising avenue for future research.

2.1   Prior Specification

We follow Frühwirth-Schnatter and Wagner (2006), Frühwirth-Schnatter and Frühwirth (2007) and Kaufmann (2015) and implement an approximate, yet accurate, MCMC estimation scheme for the unknown parameters, the latent state, and the time-varying transition probabilities. For each industry, the slope parameters of the p determinants of the intensity rate are gathered into the vector $\boldsymbol{\beta}^{i,\prime} = \left(\boldsymbol{\beta}_1^{i,\prime}, \ldots, \boldsymbol{\beta}_K^{i,\prime} \right)$. The parameters governing the transition probabilities are denoted by $\boldsymbol{\alpha}_i = \{ \boldsymbol{\alpha}_{ik} | k \in \mathcal{K}_{-k_0} \}$ with k_0 being the reference state, and $\boldsymbol{\alpha}_{ik} = (\boldsymbol{\alpha}_{i1k}^z, \ldots, \boldsymbol{\alpha}_{iKk}^z, \alpha_{i1k}, \ldots, \alpha_{iKk})$.

For the Bayesian inference to work, we need to specify the prior distributions for the model parameters. For a given state $S_{it} = k$, the prior structure is conjugate as the function for the (log of) intensity rate reduces to standard multiple linear regressions. For each state, we assume a normal

conjugate prior:

$$p\left(\boldsymbol{\beta}^i\right) = \prod_{k=1}^{K} p\left(\boldsymbol{\beta}_k^i\right) = \prod_{k=1}^{K} \mathcal{N}_p\left(b_{ik}, B_{ik}\right), \qquad \text{for} \qquad i = 1, \dots, n \qquad (4)$$

with b_{ik} and B_{ik} as the location and scale hyper-parameters for the ith industry, respectively. The prior for the βs is independent of any assumption about the transition probabilities of the states. We assume we do not have prior information about the betas and use uninformative prior hyper-parameters, i.e. $b_{ik} = \mathbf{0}$ and $B_{ik} = I \cdot 1000$, with I a $p \times p$ identity matrix. The prior specification (4) assumes the slope parameters for the intensity rates to be independent across industries and states. This makes the prior invariant to state permutations.

As far as the transition probabilities are concerned, the multi-Logit specification allows us to specify a Gaussian prior distribution for the parameter vector α_{ik}:

$$p\left(\boldsymbol{\alpha}_i\right) = \prod_{k \in \mathcal{K}_{-k_0}} p\left(\boldsymbol{\alpha}_{ik}\right) = \prod_{k \in \mathcal{K}_{-k_0}} \mathcal{N}\left(a_{ik}, A_{ik}\right), \qquad \text{for} \qquad i = 1, \dots, n \qquad (5)$$

in which, again, the independence across states makes the prior invariant to state permutations. The independence across industries and regimes allows us to specify prior hyper-parameters independently such that one prior can be imposed to be more informative for a state of, say, a merger wave. Throughout the paper we take an agnostic approach and impose the same degree of non-informativeness across states. That is, we assume we do not have prior information about the driving factors of the hidden states, i.e. $a_{ik} = \mathbf{0}$ and $A_{ik} = I \cdot 1000$, with I a $p \times p$ identity matrix.

The prior distribution for the initial state is assumed to be discrete uniform, i.e. $p\left(S_{i0}\right) = 1/K$. We complete the model by assuming that the hyper-parameters of the random-effects follow an independent Normal-Inverse Wishart conjugate prior distribution $\boldsymbol{\mu} \sim N\left(\mu_0, C_0\right)$, and $D \sim IW\left(\nu_0, D_0\right)$ with non-informative prior hyper-parameters, i.e. $\mu_0 = \mathbf{0}, C_0 = I \cdot 1000$ with I an $n \times n$ identity matrix, and $\nu_0 = n, D_0 = I \cdot 1000$.

2.2 Posterior Sampling

We follow Chib and Winkelmann (2001), Frühwirth-Schnatter and Wagner (2006), Omori et al. (2007), Frühwirth-Schnatter and Frühwirth (2007) and Kaufmann (2015) and propose an approximate, yet

accurate, Gibbs sampling scheme for both the unknown parameters and the hidden states of merger activity. This is achieved by introducing sequences of auxiliary latent variables through data augmentation. Appendix A shows the convergence properties of the Gibbs sampler and provides evidence that our model appears to be reasonably accurate when we base posterior inference on 50,000 draws with a burn-in of 10,000 and a thinning value of 10.

Let us denote with $\mathbf{y}_{i\tau:t} = (\mathbf{y}_{i\tau}, \ldots, \mathbf{y}_{it})$, $\tau \leq t$, a collection of vectors $\mathbf{y}_{iu}$ of deal flow for industry i at time u. The set of state-dependent parameters for the ith industry is defined as $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iK})$, where $\theta_{ik} = (\boldsymbol{\beta}_k^i, \boldsymbol{\alpha}_{ik}, \mu_i)$. At each iteration the Gibbs sampler sequentially cycles through the following main steps:

Step 1. Draw $S_{i1:T}$ conditional on $\boldsymbol{\theta}_i$, $\mathbf{y}_{i1:T}$ and $\mathbf{z}_{i1:T}$, for $i = 1, \ldots, n$,

Step 2. Draw the parameters $\boldsymbol{\alpha}_i$ conditional on $S_{i1:T}$ and $\mathbf{z}_{i1:T}$, for $i = 1, \ldots, n$,

Step 3. Draw $\boldsymbol{\beta}^i$ conditional on $\mathbf{y}_{i1:T}$, $\mathbf{z}_{i1:T}$, and $S_{i1:T}$, for $i = 1, \ldots, n$,

Step 4. Draw the parameters μ_i conditional on $S_{i1:T}$, $\boldsymbol{\beta}^i$, $\mathbf{y}_{i1:T}$ and $\mathbf{z}_{i1:T}$, for $i = 1, \ldots, n$.

Below, we provide a detailed description of each step of the Gibbs sampler. For ease of exposition in the notation we do not report the subscript i for each industry unless needed.

Step 1. Sampling the Latent States $S_{1:T}$

The relationship between the intensity rate and the latent state of merger activity is non-linear and non-Gaussian. In the following, we introduce two auxiliary sequences of latent processes that allow us to eliminate both the non-linearity and the non-normality and implement a standard Bayesian updating scheme.

For each $t = 1, \ldots, T$ the distribution of $y_t | \lambda_t$ may be regarded as the distribution of the number of jumps of an unobserved Poisson process with intensity λ_t having occurred in the time period $[0, \ 1]$. The first auxiliary step of our MCMC scheme follows Frühwirth-Schnatter and Wagner (2006) and creates such a Poisson process for y_t and introduces the inter-arrival times τ_{tj}, for $j = 1, \ldots, (y_t + 1)$, as missing data. We start from the fact that, conditional on the latent states and the parameters, in a

Poisson process, inter-arrival times are distributed as Exponential random variables, $\tau_{tj} \sim Exp(\lambda_t)$,

$$\tau_{tj}|\boldsymbol{\theta}, S_t = \frac{\xi_{tj}}{\lambda_t}, \qquad \xi_{tj} \sim Exp(1) \tag{6}$$

This can be rewritten as a linear non-Normal model of the form

$$-\log \tau_{tj}|\boldsymbol{\theta}, S_t = \mathbf{z}_t'\boldsymbol{\beta}_{S_t} + \mu_i + \epsilon_{tj}, \qquad \epsilon_{tj} = -\log \xi_{tj}, \qquad \xi_{tj} \sim Exp(1) \tag{7}$$

Now, let $\boldsymbol{\tau} = (\tau_{tj}, j = 1, \ldots, (y_t + 1), t = 1, \ldots, T)$ denote the collection of all the inter-arrival times. The full conditional distribution $p(S_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}_{1:T})$ depends on $\mathbf{y}_{1:T}$ only through $\boldsymbol{\tau}$, i.e. $p(S_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}_{1:T}) = p(S_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau})$. Given $\boldsymbol{\theta}, S_{1:T}$ and $\mathbf{y}_{1:T}$, the inter-arrival times are i.i.d;

$$p(\boldsymbol{\tau}|\boldsymbol{\theta}, S_{1:T}, \mathbf{y}_{1:T}) = \prod_{t=1}^{T} p\left(\tau_{t1}, \ldots, \tau_{t,y_t+1}|y_t, \boldsymbol{\theta}, S_{1:T}\right) \tag{8}$$

For fixed t, the inter-arrival times $\tau_{t1}, \ldots, \tau_{t,N+1}$, where $N = y_t$ are stochastically dependent, and the joint distribution factorizes as

$$p\left(\tau_{t1}, \ldots, \tau_{t,N}, \tau_{t,N+1}|y_t = N, \boldsymbol{\theta}, S_{1:T}\right) = p\left(\tau_{t,n+1}|y_t = N, \boldsymbol{\theta}, S_{1:T}, \tau_{t1}, \ldots, \tau_{t,N}\right) p\left(\tau_{t1}, \ldots, \tau_{t,N}|y_t = N\right)$$

That is, the first N inter-arrival times are independent of all model parameters, and are determined only by the observed number of counts y_t. By standard properties of the Poisson process, the first N arrival times are distributed as the order statistics of N uniform $[0, 1]$ random variables. Therefore, if $y_t > 0$, the joint distribution $p(\tau_{t1}, \ldots, \tau_{t,n}|y_t = N)$ is approximated by sampling the order statistics $u_{t(1)}, \ldots, u_{t(N)}$ of $N = y_t$ $Uniform[0, 1]$ random variables, and defining the inter-arrival times as their increments $\tau_{tj} = u_{t(j)} - u_{t(j-1)}$, for $j = 1, \ldots, N$, where $u_{t(0)} = 0$. Only the final inter-arrival time $\tau_{t,N+1}$ depends on the states $\mathbf{S}_{1:T}$ and the model parameters $\boldsymbol{\theta}$ through the intensity rate λ_t. Conditional on y_t, only $y_t = N$ arrivals occur in $[0, 1]$, and arrival $(N + 1)$ is known to occur after 1. Since increments are independent, the waiting time after 1 is then distributed as $Exp(\lambda_t)$. Therefore, the final arrival time is sampled from $p(\tau_{t,N+1}|y_t = N, \boldsymbol{\theta}, S_{1:T}, \tau_{t1}, \ldots, \tau_{t,n})$ by defining $\tau_{t,N+1} = 1 - \sum_{j=1}^{N} \tau_{tj} + \xi_t$, where $\xi_t \sim Exp(\lambda_t)$. The linear approximation of (7) generates a non-

Normal error term, the density of which can be approximated by a mixture of R normal components:

$$p\left(\epsilon\right) \approx \sum_{r=1}^{R} \omega_r N\left(\epsilon; m_r, \sigma_r^2\right), \tag{9}$$

where, for $r = 1, \ldots, R$, m_r and σ_r^2 are the mean and variance of a Gaussian density. We use $R = 10$ deterministic components with values ω_r, m_r, and σ_r^2 indicated in Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter and Frühwirth (2007).

Now, let $\mathbf{R} = (r_{tj}, j = 1, \ldots, (y_t + 1), t = 1, \ldots, T)$ denote the collection of latent component indicators. Conditional on $\boldsymbol{\tau}$ and $\mathbf{R}$, the non-normal, non-linear model (1) reduces to a linear, Gaussian model of the form

$$-\log \tau_{tj} | \boldsymbol{\theta}, S_t, r_{tj} = \mathbf{z}_t' \boldsymbol{\beta}_{S_t} + \mu_i + m_{r_{tj}} + \epsilon_{tj}, \qquad \epsilon_{tj} | r_{tj} \sim N\left(0, \sigma_{r_{tj}}^2\right) \tag{10}$$

To sample $\mathbf{R}$ we exploit the fact that all indicators are independent given $\boldsymbol{\tau}, \boldsymbol{\theta}, S_{1:T}$ and $\mathbf{y}_{1:T}$, i.e.,

$$p\left(\mathbf{R} | \boldsymbol{\tau}, \boldsymbol{\theta}, S_{1:T}, \mathbf{y}_{1:T}\right) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} p\left(r_{tj} | \tau_{tj}, \boldsymbol{\theta}, S_t\right) \tag{11}$$

That is, for each $t = 1, \ldots, T$ and each $j = 1, \ldots, y_t + 1$, the indicator r_{tj} is sampled independently from $p\left(r_{tj} = r | \tau_{tj}, \boldsymbol{\theta}, s_t\right)$. This density depends on the data only through τ_{tj}:

$$p\left(r_{tj} = r | \tau_{tj}, \boldsymbol{\theta}, S_t\right) \propto p\left(\tau_{tj} | r_{tj} = r, \boldsymbol{\theta}, S_t\right) \omega_r \tag{12}$$

where

$$p\left(\tau_{tj} | r_{tj} = r, \boldsymbol{\theta}, S_t\right) \propto \frac{1}{\sigma_r} \exp\left(-\frac{1}{2\sigma_r^2}\left(-\log \tau_{tj} - \mathbf{z}_t' \boldsymbol{\beta}_{S_t} - \mu_i - m_r\right)^2\right) \tag{13}$$

Finally, conditional on the hidden inter-arrival times $\boldsymbol{\tau}$, the mixture component indicators $\mathbf{R}$, and the parameters $\boldsymbol{\theta}$, the observation in (7) takes the form

$$-\log \tau_{tj} | \boldsymbol{\theta}, S_t, r_{tj} = \mathbf{z}_t' \boldsymbol{\beta}_{S_t} + \mu_i + m_{r_{tj}} + \epsilon_{tj}, \qquad \epsilon_{tj} | r_{tj} \sim N\left(0, \sigma_{r_{tj}}^2\right) \tag{14}$$

For each time t we can redefine the observation vector $\tilde{\mathbf{y}}_t$ of dimension $N = y_t + 1$ as

$$\tilde{\mathbf{y}}_t = \begin{pmatrix} -\log \tau_{t1} - \mu_i - m_{r_{t1}} \\ \vdots \\ -\log \tau_{tN} - \mu_i - m_{r_{tN}} \end{pmatrix} \tag{15}$$

That is, we can rewrite an augmented model in a state-space form as

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{z}}_t' \boldsymbol{\beta}_{S_t} + \eta_t, \qquad \eta_t \sim N\left(0, \Sigma_t\right), \tag{16}$$

where $\Sigma_t = diag\left(\sigma_{r_{t1}}^2, \ldots, \sigma_{r_{tN}}^2\right)$ and $\tilde{\mathbf{z}}_t$ contains N rows of $\mathbf{z}_t$, i.e., $\tilde{\mathbf{z}}_t = (\mathbf{z}_t', \ldots, \mathbf{z}_t')'$.

Now we have a state-space model with a repeated measurement in which the transition equation for the hidden states $S_{1:T}$ is the same as for the original Poisson regression model (1). As a matter of fact, (16) can be treated as a Markov regime-switching seemingly unrelated regression model. As a result, we can now implement a standard Forward Filtering Backward Sampling (FFBS) algorithm for discrete latent states by iterating two steps: the prediction step at each time t

$$p\left(S_{t+1} = j | \boldsymbol{\theta}, \tilde{\mathbf{y}}_{1:t}\right) = \sum_{k=1}^{K} \pi_{kj,t} p\left(S_t = k | \boldsymbol{\theta}, \tilde{\mathbf{y}}_{1:t}\right) \tag{17}$$

and the subsequent updating step

$$p(S_{t+1} = k | \boldsymbol{\theta}, \tilde{\mathbf{y}}_{1:t+1}) = \frac{p(\tilde{\mathbf{y}}_{t+1} | S_{t+1} = k, \boldsymbol{\theta}, \tilde{\mathbf{z}}_t) p(S_{t+1} = k | \tilde{\mathbf{y}}_{1:t}, \boldsymbol{\theta})}{p\left(\tilde{\mathbf{y}}_{t+1} |, \tilde{\mathbf{z}}_t, \boldsymbol{\theta}\right)} \tag{18}$$

where $p(\tilde{\mathbf{y}}_{t+1} | S_{t+1} = k, \boldsymbol{\theta}, \tilde{\mathbf{z}}_t) = N\left(\tilde{\mathbf{z}}_t' \boldsymbol{\beta}_k, \Sigma_{t+1}\right)$, and the normalising constant is the marginal predictive likelihood defined as

$$p\left(\tilde{\mathbf{y}}_{t+1} | \tilde{\mathbf{z}}_t, \boldsymbol{\theta}\right) = \sum_{k=1}^{K} p\left(\tilde{\mathbf{y}}_{t+1} | S_{t+1} = k, \boldsymbol{\theta}, \tilde{\mathbf{z}}_t\right) p\left(S_{t+1} = k | \boldsymbol{\theta}, \tilde{\mathbf{y}}_{1:t}\right) \tag{19}$$

Posterior draws of the latent states can then be obtained recursively and backward in time by using the smoothed probabilities, for example,

$$p\left(S_t = k | S_{t+1} = j, \mathbf{y}_{1:t}, \boldsymbol{\theta}\right) = \frac{\pi_{kj,t} p\left(S_t = k | \mathbf{y}_{1:t}, \boldsymbol{\theta}\right)}{p\left(S_{t+1} = j | \theta, \mathbf{y}_{1:t}\right)} \tag{20}$$

Step 2. Sampling the Parameters $\boldsymbol{\alpha}_k$

The sampling of the parameters governing the time-varying transition probabilities $\pi_{lk,t}$ are drawn based on a data augmentation process as in Kaufmann (2015). In a first step, we extend the model to a non-normal specification for so-called state-dependent latent utilities:

$$S_{k,t}^u = \mathbf{Z}'_{t-1}\boldsymbol{\alpha}_k + \nu_{k,t}, \qquad \forall k \in \mathcal{K}_{-k_0},$$

$$S_{k_0,t}^u = \nu_{k_0,t}, \quad \text{for the identification restriction of the reference state} \quad \alpha_{k_0} = 0, \tag{21}$$

where $Z_t = \left(\mathbf{z}_t D_t^{(1)}, \mathbf{z}_t D_t^{(2)}, \dots, \mathbf{z}_t D_t^{(K)}, D_t^{(1)}, D_t^{(2)}, \dots, D_t^{(K)}\right)$ with $D_t^{(k)}$ being a dummy variable that takes a value equal to one if $S_t = k$. The errors $\nu_{k,t}$ are i.i.d. and follow a type I extreme value distribution. Conditional on the latent state variables $S_{1:T}$, we sample $S_{k,t}^u$ by using a set of independent random draws $V_{1t}, \dots, V_{Kt}$ from a $Uniform[0, 1]$, such that

$$S_{k,t}^u = -\log\left(-\frac{\log\left(V_{1t}\right)}{\sum_{k=1}^K \tilde{\lambda}_{kt}} - \frac{\log\left(V_{kt}\right)}{\tilde{\lambda}_{kt}}\mathbb{I}_{\{S_t \neq k\}}\right), \tag{22}$$

with $\tilde{\lambda}_{kt} = \exp\left(\mathbf{Z}'_{t-1}\boldsymbol{\alpha}_k\right)$. The simulation step (22) is derived by exploiting the assumption that the maximal value of $S_{j,t}^u$ is obtained in correspondence of the observed state. That is, $\exp\left(-S_{j,t}^u\right)$ is the minimum value among all of the possible alternatives $\exp\left(-S_{k,t}^u\right)$ if $S_t = j$. The type I extreme value distribution of $\nu_{k,t}$ implies that $\exp\left(-S_{k,t}^u\right) \sim Exp\left(\tilde{\lambda}_{kt}\right)$.

Given the latent utilities S_{kt}^u, we can now sample the parameters $\boldsymbol{\alpha}_k$. However, the dynamics implied by (21) is non-Gaussian. Just as in Step 1 above, we can approximate such distribution by using a mixture of Normal distributions. Let $\mathbf{R}^s = \left(r_{t,k}^s, k = 1, \dots, K, t = 1, \dots, T\right)$ denote the collection of latent component indicators. Given $S_{1:T}^u$ and $\mathbf{R}^s$, the non-normal, non-linear model (21) reduces to

$$S_{k,t}^u = \mathbf{Z}'_t\boldsymbol{\alpha}_k + \tilde{m}_{r_{t,k}^s} + \epsilon_{t,k}, \qquad \epsilon_{t,k}|r_{t,k}^s \sim N\left(0, \tilde{\sigma}_{r_{t,k}^s}^2\right)$$

To sample $\mathbf{R}^s$ we exploit the fact that all indicators are conditionally independent given the state-utilities. As such, the density $p\left(\mathbf{R}^s|S_{1:T}^u, \boldsymbol{\alpha}, \mathbf{z}_{1:T}\right)$ depends on the observable data only through $\mathbf{R}^s$:

$$p\left(r_{t,k}^s = r|S_{k,t}^u, \boldsymbol{\alpha}_k\right) \propto p\left(S_{k,t}^u|r_{t,k}^s = r, \boldsymbol{\alpha}_k, S_{1:T}\right)\omega_r^s,$$

12

with

$$p\left(S_{k,t}^u | r_{t,k}^s = r, \boldsymbol{\alpha}_k, s_{1:T}\right) \propto \frac{1}{\tilde{\sigma}_r^2} \exp\left(-\frac{1}{2\tilde{\sigma}_r^2}\left(S_{k,t}^u - \mathbf{Z}_{t-1}'\boldsymbol{\alpha}_k - \tilde{m}_r\right)^2\right), \qquad k \in \mathcal{K}_{-k_0},$$

where $r = 1, \ldots, 10$ and the respective $\tilde{m}_r$, $\tilde{\sigma}_r^2$ and $\tilde{\omega}_r$ are deterministic values taken from Frühwirth-Schnatter and Frühwirth (2007). Notice that the Gaussian mixture approximations for $\mathbf{R}$ and $\mathbf{R}^s$ are drawn separately.

Finally, for all given state-utilities $S_{1:T}^u = \left(S_{1,1}^u, \ldots, S_{K,1}^u, \ldots, S_{K,T}^u\right)$ and all component indicators $\mathbf{R}^s = (r_{11}^s, \ldots, r_{K1}^s, \ldots, r_{KT}^s)$, we obtain a standard linear regression model for the parameters governing the transition probabilities to each state, except for the reference state k_0, which implies $\alpha_{k_0} = 0$. Given the conjugate prior structure (5), the posterior distribution is updated as

$$\left(\boldsymbol{\alpha}_k | \mathbf{y}_{1:T}, S_{1:T}^u, \mathbf{R}^s\right) \sim \mathcal{N}_N\left(A_k^*\left(\sum_{t=1}^T \mathbf{Z}_{t-1}'\left(S_{k,t}^u - \tilde{m}_{r_{t,k}^s}\right)/\tilde{\sigma}_{r_{t,k}^s}^2 + A_k^{-1}a_k\right), A_k^*\right), \qquad (23)$$

with $A_k^* = \left(\sum_{t=1}^T \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'/\tilde{\sigma}_{r_{t,k}^s}^2 + A_k^{-1}\right)^{-1}$ being the posterior scale parameter.

Step 3. Sampling the Slope Parameters $\boldsymbol{\beta}_k$

Conditional on $S_{1:T}, \mathbf{y}_{1:T}$ and the Gaussian linear approximation (16), the prior (4) is conjugate; that is, we can use standard Bayesian updating rules for a multivariate Normal distribution. Let $\mathcal{T}_k = \{t : s_t = k\}$ the sample conditional on the state k, the posterior for the regime-dependent betas $\boldsymbol{\beta}_k$ is defined as

$$\left(\boldsymbol{\beta}_k | \mathbf{y}_{1:T}, S_{1:T}\right) \sim \mathcal{N}_p\left(B_k^*\left(\sum_{t \in \mathcal{T}_k} \tilde{\mathbf{z}}_t'\tilde{\mathbf{y}}_t + B_k^{-1}b_k\right), B_k^*\right) \qquad (24)$$

with $B_k^* = \left(\sum_{t \in \mathcal{T}_k} \tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t' + B_k^{-1}\right)^{-1}$ being the posterior scale parameter. Notice that the conditional covariance Σ_t is given by the latent indicators $\mathbf{R}$.

Step 4. Sampling the Random-Effects

We follow Chib and Winkelmann (2001) and sample the random-effects via a Metropolis-Hastings (MH) step with a random-walk proposal. That is, given the current value of the parameter μ_i, a

sample proposal value μ_i^* is generated from $q\left(\mu_i, \mu_i^*\right) = p\left(\mu_i^* | \mu_i, \tilde{\mu}D\right)$ where $\tilde{\mu}$ is a scalar that is adjusted in trial runs to obtain suitable candidates. The acceptance rate probability is calculated as $\psi\left(\mu_i, \mu_i^*\right) = \min\left(p\left(\mu_i^*\right) q\left(\mu_i^*, \mu_i\right) / p\left(\mu_i\right) q\left(\mu_i, \mu_i^*\right), 1\right)$ where

$$p\left(\mu_i\right) = N\left(\mu_i | \boldsymbol{\mu}, D\right) \prod_{t=1}^{T} \exp\left[-\exp\left(\mathbf{z}_{it}'\boldsymbol{\beta}_{S_t} + \mu_i\right)\right] \times \left[\exp\left(\mathbf{z}_{it}'\boldsymbol{\beta}_{S_t} + \mu_i\right)\right]^{y_{it}}$$

Once a cycle of the Metropolis-within-Gibbs sampler is completed for each industry, we can estimate the hyper-parameters that drive the random-effects. The Normal-Inverse Wishart independent prior in Section 2.1 is conjugate, which means that posterior draws can be sampled from the posterior distribution

$$\boldsymbol{\mu}|D,\ldots \sim \mathcal{N}\left(C^*\left(C_0^{-1}\mu_0 + \sum_{i=1}^{n} D^{-1}\mu_i\right)^{-1}, C^*\right) \qquad \text{with} \qquad C^* = \left(C_0^{-1} + nD^{-1}\right)^{-1},$$

and D from

$$D|\boldsymbol{\mu},\ldots \sim IW\left(n + \nu_0, \left[D_0 + \sum_{i=1}^{n} \left(\mu_i - \boldsymbol{\mu}\right)\left(\mu_i - \boldsymbol{\mu}\right)'\right]^{-1}\right)$$

One comment is in order. The choice of a random-walk proposal for the MH step arguably generates a high correlation in the sample of draws from the Markov chain. To increase the effective sample size, we keep one in ten draws and discard the draws in between. This substantially reduces the amount of autocorrelation in the posterior sampling. Appendix A reports a set of convergence diagnostics that show the good performance of our algorithmic procedure.

3 Data Description

The sample contains all announced bids for US private and public acquirers that were announced in the period from 1983 to 2016, for which the bidder did not previously own a majority interest in the target and was seeking to obtain a majority interest through the transaction. Data on M&A deal flow are collected from Thomson One Banker and complemented with firm-level stock market and accounting data from the Centre for Research in Security Prices (CRSP) and Compustat databases, respectively. Macroeconomic factors are collected from the Federal Reserve Economic Data series and from the

official providers of the relevant indexes. A deal is included in the final sample if the transaction value is above $5 million and the transaction is not a buyback, an exchange offer, a recapitalization, an acquisition of partial or remaining interest, a spin-off, a self-tender or a repurchase. The sample includes 65,073 observations overall, for which deal value was disclosed.

M&A deals are aggregated at the industry level for each quarter, assigning each deal to one of twelve industries based on the four-digit SIC code of the bidder at the time of the announcement. We use the twelve industry classification codes obtained from Kenneth French's website. We form a panel by matching the number of deals observed in each quarter and for each industry with a set of variables proxying for economic shocks and market valuations.

We follow Harford (2005) and construct industry-specific economic shocks as the first principal component of the median absolute annual changes of a set of industry-specific fundamentals variables: margin on sales (i.e. Margin, the ratio of net income over sales), asset turnover (i.e. AT, the ratio of sales over assets at the beginning of the period), research and development (i.e. R&D expense scaled by assets at the beginning of the period), capital expenditures (i.e. CAPEX, scaled by assets at the beginning of the period), growth in the number of employees (Emp), return on assets (ROA), and sales growth (Sale). These variables are typically associated to the neoclassical hypothesis on merger waves (see, for example, Coase, 1937, Gort, 1969, Shleifer and Vishny, 1992, Harford, 1999, Andrade et al., 2001, Holmstrom and Kaplan, 2001, Maksimovic and Phillips, 2001, Jovanovic and Rousseau, 2002, and Harford, 2005). All these variables are computed by aggregating firm-level data available on Compustat in each industry, such that we allow economic shocks to have different effects and magnitudes across industries. In addition, other factors related to the neoclassical hypothesis are contemporaneous capital liquidity (C&I), measured in terms of the spread between the average interest rate on Commercial and Industrial loans, and the Fed Funds rate, as published by the Federal Reserve.

The panel consists also of variables that capture market timing and valuations, such as the industry-specific average book-to-market ratio (BM), its cross-sectional standard deviation (std(BM), computed across all firms with data available on Compustat for each industry), and value-weighted industry stock returns (Returns). These variables are often associated to the behavioural hypothesis outlined in Shleifer and Vishny (2003), Rhodes-Kropf and Viswanathan (2004), and Rhodes-Kropf et al. (2005), among others.

We also consider a set of variables that capture aggregate macroeconomic conditions (see, for example, Melicher et al., 1983, Shugart and Tollison, 1984, Becketti, 1986, Town, 1992, Golbe and White, 1993, Mulherin and Boone, 2000, Andrade et al., 2001, and Choi and Jeon, 2011). These include: year-on-year monthly compounded industrial production growth as a measure of changes in aggregate output, the credit spread (measured as the yield spread between Moody's 20-year Baa- and Aaa-rated corporate bonds), and the real risk-free rate (proxied by the difference between the one-month T-bill rate and the CPI log inflation rate). Table 1 provides a detailed description of sources and frequencies of the determinants investigated in the empirical analysis.

[Insert Table 1 about here]

Merger deals are clustered at quarterly frequency as a trade-off between having enough granularity in the data and keeping enough information; namely, this means there are a sufficient number of deals for each data point and for each industry. In order to ensure a coherent comparison across sets of predictors, we collected the data on both financial variables and economic shocks at the quarterly frequency, as well.

Shocks to an industry environment can also come from major regulatory changes. We define a dummy variable to indicate whether or not the industry has recently been subject to one of the deregulatory events.

[Insert Table 2 about here]

Table 2 documents major deregulatory initiatives during the sample period 1983:Q1-2016:Q4. Deregulatory events are considered at the industry level, assigning each event to one of twelve industries based on the SIC code of the bidder at the time of the announcement. The events are constructed from Viscusi et al. (2005) and by searching for recent major initiatives in Factiva.

4   Empirical Results

Following Harford (2005), we call the first principal component from the seven economic shock variables (Margin, AT, R&D, CAPEX, Emp, Roa, and Sales) the Economic Activity index. This index is included in the regression together with the regressors justified by the behavioral theory, a dummy for regulatory changes, the proxy for capital liquidity and the set of macro-financial factors, which

makes $\mathbf{z}_t$ composed of nine explanatory variables.

For the ease of exposition, we report the time series of deals for four representative industries, namely Manufacturing, Telecommunications, Money and Healthcare. Although this choice looks restrictive, it turns out that these four sectors are fairly representative of overall market activity. Figure 1 shows the unconditional distribution of the deal flow across industries, calculated on the basis of the average number of deals over the sample period for a given industry relative to the average number of deals at the aggregate market level.

[Insert Figure 1 about here]

The figure makes clear that Money represents the biggest market where M&A is implemented and that, together with Manufacturing, Telecommunications and Healthcare, it represents almost half of the total market activity. Figure 2 shows the time series of market transaction data (light blue bars).

[Insert Figure 2 about here]

The dynamics of deal flow is not synchronous across industries but varies in the cross-section. For instance, while the deal flow for Money $y_{Money,t}$ (bottom-left panel) peaks in the second half of the 90s and from 2005 to 2010, the intensity of M&A activity for the Telecommunications sector is sensibly increasing from the late 80s, while it is substantially flat from the beginning of the 2000s. Also, different industries show different persistence in periods of high merger activity. For instance, the deal flow is remarkably high for almost ten years for the Telecommunications industry, while it lasts a few years for Manufacturing and Other. Notably, across all industries, the major M&A activity is concentrated between the late 90s and the first half of the 2000s.

We first test the significance of determinants of merger waves for each industry separately. In the modeling framework, we explicitly consider industry-specific random-effects, which are drawn from a common cross-sectional distribution. We assume throughout that two regimes are enough to characterize the dynamics of merger waves in the time series of deal flow. In Section 4.2, we formally assess out-of-sample the assumption of two regimes by calculating the log-predictive density ratios of a model with three regimes and a model with two regimes. The results show that a two-state specification compares favorably to the three-state alternative specification.

4.1 The Dynamics of Merger Waves

In terms of the two states, we label $S_t = 1$ as "no-wave" and $S_t = 2$ as "wave", respectively. In our modeling setting, the significance of a specific factor to determine a change to a state of merger wave can be directly tested on the basis of posterior estimates of $\boldsymbol{\alpha}_{12}^z = \left(\boldsymbol{\alpha}_{12,1}^z, \ldots, \boldsymbol{\alpha}_{12,N}^z \right)$. For instance, a positive and significant $\boldsymbol{\alpha}_{12,1}^z$ implies that the first factor can explain a regime switch towards a merger wave. On the other hand, if $\boldsymbol{\alpha}_{12,1}^z$ is not statistically different from zero, it means that the first factor does not significantly affect regimes of merger activity. Figure 3 shows the posterior distributions of $\boldsymbol{\alpha}_{12}^z$ across four representative industries, namely Manufacturing, Telecommunications, Money and Healthcare. Posterior estimates are obtained from the Gibbs sampler detailed in Section 2. For ease of exposition, we do not report the posterior estimates of those parameters that are not statistically significant at the 95% confidence level, i.e. those where the zero value enters the 95% posterior credibility interval.

[Insert Figure 3 about here]

The cross-sectional standard deviation of the book-to-market ratio significantly and negatively affects the dynamics of merger waves, with the exception of the financial sector. This is consistent with the evidence provided by Shleifer and Vishny (2003), Rhodes-Kropf and Viswanathan (2004), and Rhodes-Kropf et al. (2005), which showed that the increasing volatility of valuation ratios can have a negative effect on merger activity, meaning that when market volatility increases the aggregate propensity to engage in M&A decreases. Similarly, the level of the book-to-market ratio represents a significant predictor of aggregate M&A activity for the financial sector and the Telecommunications industry. Again, this is consistent with the existing evidence and the behavioral hypothesis, which posits that merger waves tend to occur when market prices are higher than fundamentals, i.e. market timing. This explains the negative relationship between the book-to-market ratios and merger waves. On the other hand, past returns do not seem to be significantly related to the one-step ahead probability of being in a state of merger wave.

The impact of economic activity on the dynamics of merger waves is null except for in the Money sector. Overall, shocks to industry-specific economic conditions do not represent significant driving factors of transitions towards periods of abnormally high merger activity. As a result, the empirical evidence does not support the neoclassical theory on merger waves for most industries. Interestingly,

and partly consistent with the results of Harford (2005), capital liquidity negatively affects merger waves for the Money industry. As liquidity conditions gets tighter, the propensity to engage in merging deals in the financial sector decreases.

One comment is in order. We do not argue that capital liquidity per se does not have a direct effect on merger activity. Indeed, non-financial firms could be heavily affected by the access to credit markets. Instead, we assess that the explanatory power of liquidity conditions can be mitigated once including the aggregate business cycle in the regression specification. Indeed, aggregate economic and credit conditions are closely related. This intuition is captured by the coefficient on industrial production for the non-financial industries, whose interpretation should be opposite to the one on capital liquidity, as proxied by the C&I loan spread. Output growth positively and significantly affects the majority of industries, i.e. $\boldsymbol{\alpha}_{12}^z \neq 0$. Regulatory changes turn out to increase the propensity of engaging in M&A for Telecommumications. This is consistent with the results provided in Harford (2005).

The posterior estimates of the sensitivity parameters in Figure 3 suggest that macroeconomic fundamentals play a significant role in the dynamics of merger waves, supporting a more general "merger activity-economic prosperity" theory to complement the behavioral explanation (see, for example, Reid, 1968, Melicher et al., 1983, Shugart and Tollison, 1984, Becketti, 1986, Town, 1992, Golbe and White, 1993, Mulherin and Boone, 2000, Andrade et al., 2001, and Choi and Jeon, 2011 for related discussions). As a matter of fact, the slope on industrial production turns out to be a significant driving force for the state of merger wave. Interestingly, the effect of economic growth is opposite for financial vs non-financial firms, with output growth that is negatively related to M&A activity for the latter. This is consistent with the conventional wisdom that posits that consolidation and re-structuring in the financial sector is often countercyclical. As far as the non-financial sectors are concerned, higher output growth leads to an increased probability of being in a merger wave. In other words, for non-financial sectors, merger waves tend to be positively correlated with the business cycle.

The ability to capture the heterogeneity in the significance of merger wave determinants is a key feature of our model. In fact, by aggregating deals at the market level, one would likely average out the longitudinal variation. Figure 4 shows this case in point. The left panel shows the posterior estimates of the industry-specific random-effects for representative sectors, namely Manufacturing,

Telecommunications, Healthcare, and Money.

[Insert Figure 4 about here]

The estimates make clear that there is indeed a substantial degree of heterogeneity in the industry fixed-effect, meaning that there are non-trivial effects across industries that are not directly captured by our set of explanatory variables. The right panel of Figure 4 shows the posterior estimates of the cross-sectional covariance matrix D in Step 4 of the Gibbs sampler outlined in Section 2. Although random effects are highly significant, the low cross-sectional correlations show that they share only a mild common structure. This suggests that, conditional on the set of predictive variables, the residual co-movement across merger waves is low.

The low cross-sectional correlations for the industry random-effects suggest that the model-implied posterior estimates of the probabilities of being in a state of high merger activity can be quite different across industries. Figure 5 shows that this is indeed the case. The posterior, i.e. smoothed, merger wave probability changes quite substantially in the cross-section.

[Insert Figure 5 about here]

While for the Manufacturing industry (top-left panel) abnormally high M&A activity is mostly clustered in a few years towards the end of the 90s, the posterior estimates for merger waves for Telecommunications (top-right panel), for example, tend to be more spread throughout the sample. More generally, consistent with the results in Harford (2005), most of the wave periods are located across the second half of the 90s and the beginning of the 2000s.

However, two important unexplored features emerge. First, coherent with the time series behaviour of deal flow, the length of model-implied merger waves is not homogeneous across industries, as for some industries, such as Manufacturing, for example, the length of abnormally high merger activity can be much more than for Telecommunications, Healthcare and Money. Such heterogeneity in persistence would be ignored by pooling together information on deal flow and investigating the determinants of merger waves at the aggregate level. Second, waves, although overlapping to some extent, are not contemporaneous across industries. For instance, while the merger wave for Telecommunications mostly coincided with the economic prosperity of the mid- to late-80s and the second

20

half of the 90s, the same explanation would not apply for the Healthcare and the Money industries. Differently, the merger waves in these industries would be consistent with the recovery after the economic recession of 1990–91, as confirmed by a positive effect of output growth on the propensity to engage in a deal.

Such heterogeneity is primarily due to the differences in the persistence of the transition probabilities. Figure 6 shows the posterior estimates of $\boldsymbol{\alpha}_{22}^{z}$, which are the parameters that determine the persistence of a state of high merger activity.

[Insert Figure 6 about here]

For the sake of comparison with Figure 3, we report the posterior estimates of the slope parameters for the same explanatory variables. Two results emerge. First, it is evident that the factors driving the persistence of the merger wave state are not the same across industries. For instance, while the persistence of the latent state for Manufacturing is significantly affected by valuations uncertainty, the same is not true for the Money sector, where the cross-sectional standard deviation of the book-to-market ratio does not play any significant role. Second, some of the factors that contribute in accelerating merger activity do not affect the persistence of merger waves. For instance, while the top-left panel of Figure 3 shows that C&I is highly significant for the dynamics of $\boldsymbol{\alpha}_{12}^{z}$, the same variable does not have a significant effect on the persistence of $S_t = 2$. Similarly, both C&I and the book-to-market ratio turn out to be insignificant for the persistence of merger waves (see bottom-left panel). To summarize, Figures 5 provide evidence that, despite some overlapping, merger waves are not synchronous across industries over time. More precisely, Figure6 shows that such diversity is primarily due to the heterogeneity in the significant factors that drive the time-varying transition probabilities.

5   Model Assessment

Figures 3 to 5 together support the idea that determinants of merger waves are heterogeneous in nature and specific for different industries. These empirical results ground on the idea that two regimes are enough to characterize the dynamics of merger waves in the time series of deal flow. Figure 7 gives a visual impression of the model in-sample performance. The solid red line represents the posterior

median of $\lambda_{S_t,t}$ as specified in (1), which represents an estimate of the expected number of deals. The fit of the model is quite accurate.

[Insert Figure 7 about here]

Except for the initial part of the merger wave in the Healthcare industry, posterior estimates of the expected number of deals closely follow peaks in merger activity. The model fits reasonably well the deal flow for Money (bottom-left panel), Manufacturing (top-left panel), and Telecommunications (top-right panel). As a whole, Figure 7 suggests that a non-linear, non-Gaussian model specification with two regimes is likely coherent with the time-series of deal flows.

We now compute and compare the benchmark two-state regime-switching specification against both a three-state and a one-state alternative by comparing the density forecasts; and we then evaluate the log predictive density ratios (LPDR), at horizon k and across time indices t, that is,

$$\text{LPDR}_{it}(k) = \sum_{\tau=1}^{t} \log\{p(y_{i\tau+k}|y_{i1:\tau}, \mathcal{M}_s)/p(y_{i\tau+k}|y_{i1:\tau}, \mathcal{M}_0)\}, \qquad \text{for} \qquad i = 1, \ldots, n, \qquad (25)$$

where $p(y_{i\tau+k}|y_{i1\tau}, \mathcal{M}_s)$ is the marginal predictive density computed at time τ for the horizon $\tau + k$ under the competing specification indexed by $\mathcal{M}_s$, compared to our forecasting framework labeled by $\mathcal{M}_0$. Marginal predictive densities allow us to make a robust comparison across models as they take into account the latent nature of state indicators for merger waves and the uncertain nature of the model parameters. In that respect, by calculating the marginal predictive density we provide a measure of the ability of the model to explain not only the expected value, i.e. the intensity, of merger deals, but also their overall distribution, naturally penalizing the size/complexity of different models. The marginal likelihood for each model is not available in closed form and must be approximated numerically.

We approximate the predictive density by using an importance sampling (IS) estimator (see, for example, Geweke, 2005, and Geweke and Amisano, 2012), in which sampling draws from the posterior distribution of the parameters are obtained using the entire sample of observable predictors and latent expected returns. The IS estimator is expected to work well in practice when the Gibbs sampler covers reasonably well the parameter region where the conditional likelihood is large. This is usually the case for low dimensional models like ours. In addition, conditional on using the same

set of parameters across forecasting horizons, the importance-sampling estimator is unbiased and the predictive likelihood can be consistently estimated from the posterior (see, for example, Geweke, 2005). We simulate in each Gibbs step $y_{i\tau+h}|y_{i1:\tau}$ using (1) and the Markov regime-switching dynamics, where we replace the parameters and the in-sample latent variables by the draw from the posterior distribution.

As used by several authors recently, LPDR measures provide a direct statistical assessment of relative accuracy and can be interpreted as a log predictive Bayes factor under the assumption of equal prior models' probabilities. In this respect, the LPDR weighs and compares dispersion of forecast densities along with location, giving a broader understanding of the predictive abilities of the different regime assumptions.

We obtain the one-quarter ahead predictions through direct forecasting, as indirect procedures would entail iterating forward the posterior predictive distribution multiple times, including predicting the determinants, and thus repeatedly rerunning the MCMC sampler. Also, the Markov regime-switching dynamics of the model imply that direct and indirect forecasting are approximately numerically equivalent. The forecast at time t is based on a re-estimation of the model using an expanding window of merger activity determinants. We start from an initial training sample of $\tau_0 = 60$, and produce a one-step ahead forecast $\tau_0 + 1$, which is then evaluated against the observable deal flow at time $\tau_0 + 1$. We then repeat this process until the end of the sample and compute the predictive densities to measure (25).

We compute the predictive densities both separately for each industry and jointly for the overall economy. Figures 4 and 5 show that there is a mild but significant correlation for the merger waves across industries. In this respect, the joint predictive density provides additional information about the out-performance of the two-state model vs alternatives. In order to keep the graphs readable, we rescaled the LPDR by its sample deviation. The top panels of Figure 8 show the results comparing the three-state vs the two-state specification.

[Insert Figure 8 about here]

The top-left panel shows that the out-of-sample performance is strongly in favour of the two-state regime switching specification across industries (see Kass and Raftery, 1995 for a discussion of threshold values related to the log predictive Bayes factor). Except for the Telecommunicaitons industry

in the early 2000s, the out-performance of the Poisson regression with two vs three regimes steadily increases over time, especially for those sectors more correlated with the business cycle, such as Manufacturing, Consumer Durables, Business Equipment and Consumer Non-Durables. The top-right right panel shows the joint LPDR; again, the two-state specification fits the data better.

The bottom panels of Figure 8 show the results comparing the one-state vs the two-state specification. With the only exception being Manufacturing in the late 90s, a model with two regimes vs state invariant generates a much larger marginal predictive density. Similarly, the right panel shows that the outperformance of the two-state specification is not limited to a few industries, but is confirmed by the joint marginal predictive density. As a whole, the empirical evidence provided by Figure 8 strongly points towards a model with two regimes vs both a three- and a one-state alternative.

6   Conclusion

We propose a novel Markov regime-switching Poisson regression model with time-varying transition probabilities to rationalize wave-like patterns in the intensity rate of industry-specific merger activity. Empirically, we show that merger waves vary significantly across industries, both in terms of their timing and persistence. Such differences are mainly due to a significant heterogeneity in the factor exposures. That is, the cross-sectional differences in the dynamics of merger waves can be explained by differences in the transition dynamics that drive the states of merger waves, even after accounting for cross-sectional random-effects. This suggests that any inference on existing economically justified competing explanations of merger waves would suffer from generalization at the aggregate market level, as the observed cross-industry heterogeneity in waves is shown to be the consequence of different responses to common or distinct drivers of merger activity.

Appendix

A   Convergence Diagnostics

To further to understand the properties of our modelling framework, we investigate the convergence properties of the MCMC algorithm outlined in Section 2. The results below are based on 50,000 posterior draws with a burn-in of 10,000 and a thinning value of 10. If the algorithm is designed properly, we should expect that the Markov chain of the posterior draws of the parameters will convergence to some stable distribution (see, e.g., Gelman and Rubin, 1992, Geweke, 1992, Raftery and Lewis, 1992, and Brooks and Gelman, 1998 for details).

One way to see if our Markov chain is mixing is to see how fast it convergences to the same stable mean. Th left panel of Figure A.1 shows the running mean plot. A running mean plot is a plot of the iterations against the mean of the draws up to each iteration. More precisely, Figure A.1 reports the running mean for the betas averaged across predictors for the Consumer Durable industry. Results for other industries are qualitatively the same. The benchmark specification is with $K = 2$ and with all predictors included.

[Insert Figure A.1 about here]

A good mixing chain is assumed to converge quickly to some long-run mean value. This is indeed the case for both the chain conditional on a state of high merger activity (red line) and low merger activity (blue line). The running mean quickly converges for both chains and remains highly stable after the burn-in sample (vertical black line).

Another graphical way to assess convergence is to assess the autocorrelations between the draws of the Markov chain. If autocorrelation is still relatively high for higher lags, this indicates a high correlation between draws and slow mixing. That is, the higher the autocorrelation, the lower the effective sample size used to approximate the posterior quantities of interest.

The right panel of Figure A.1 shows the autocorrelation function (ACF) for the average betas across predictors for the Consumer Durable industry. The benchmark specification is with two regimes and all regressors. The ACF is low from the very initial lags for the chain conditional on both merger wave (blue line) and no merger wave (red line). The low autocorrelation is due to the fact that we

keep one in ten draws and discard the others. This significantly reduces the persistence of the draws.

Albeit informative, the evidence provided in Figure A.1 is not conclusive. We now test the Markov chain convergence by implementing a test of difference of means to investigate if draws are from the same distribution (see Geweke, 1992). The test statistic is a standard Z-score with standard errors computed using the Newey and West (1987) heteroscedasticity and autocorrelation robust variance estimator with a bandwidth set to 4%, 8%, and 15% of the utilized sample size. Table A.1 shows the results for the average betas for the explanatory variables in the merger waves dynamics across different industries.

[Insert Table A.1 about here]

Column one of the summary of convergence diagnostics shows that even without correcting for autocorrelation the t-statistic does not reject the null hypothesis of equal means in the sub-samples for seven out of twelve industries, and only for Energy, Telecomm and Utilities the null hypothesis is rejected at the 99% confidence level. The evidence of convergence significantly increases by adjusting the standard errors for autocorrelation and heteroscedasticity. While for a bandwidth of 4% of the data only Telecomm, Utilities and Energy reject the null hypothesis at the 95% confidence level, the Z-score calculated using a bandwidth of 15% shows strong evidence of convergence with none of the industries rejecting the null hypothesis.

To summarize, Figure A.1 and Table A.1 provide evidence that our model appears to be reasonably accurate when we base posterior inference on 50,000 draws with a burn-in of 10,000 and a thinning value of 10, which keeps the computational burden relatively low.

Figure 1: Distribution of Deals

This figure shows the unconditional distribution of deals, calculated as the sample average of the deals flow relative to the aggregate number of deals across the sample. Deals are measured as the number of all M&A bids announced by US private and public acquirers for US public and private targets. We considered those deals with a value higher than $5mln (including net debt of the target), and exclude those identified as spinoffs, recapitalizations, self-tenders, exchange offers and repurchases. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. The sample period is 1983:Q1-2016:Q4.
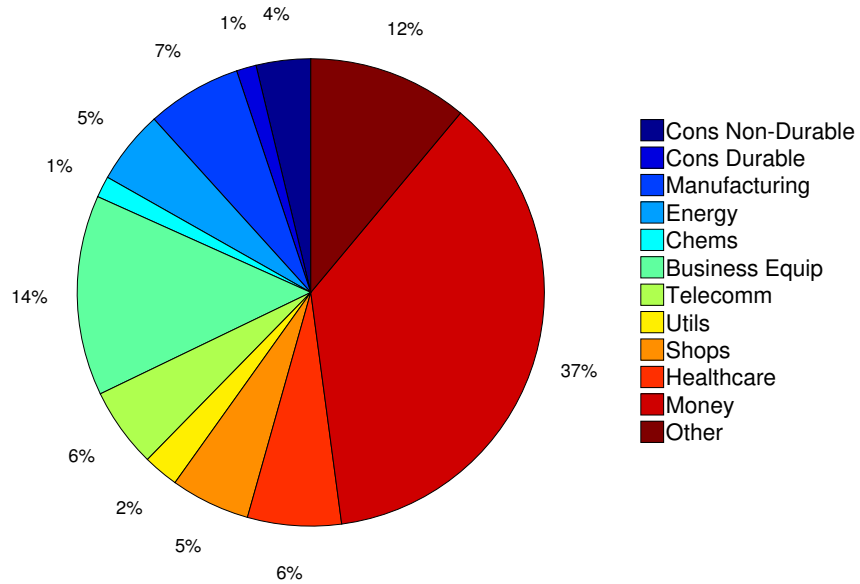
Figure 2: Time Series of Industry Merger Activity

This figure reports the time series of merger activity across industries, measured as the number of all M&A bids announced by US private and public acquirers for US public and private targets in the period 1983:Q1 to 2016:Q4. The left axis on each graph represents the number of deals.
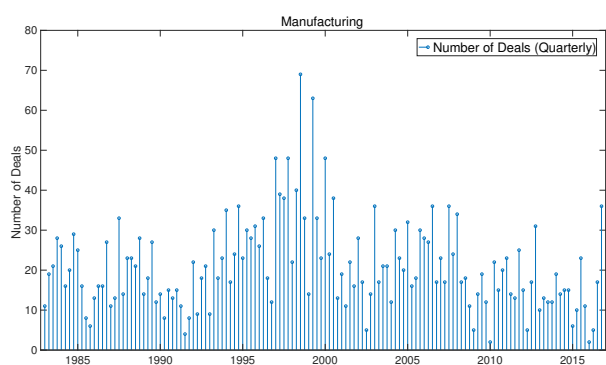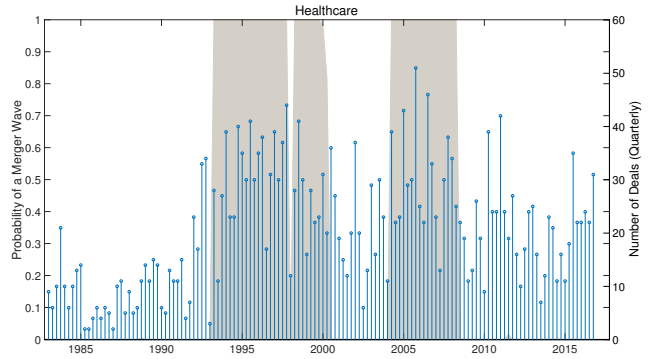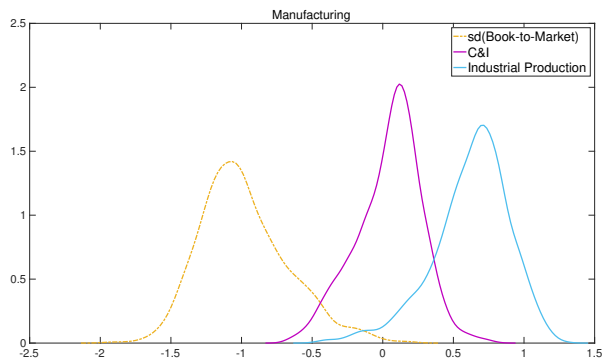

(a) Manufacturing
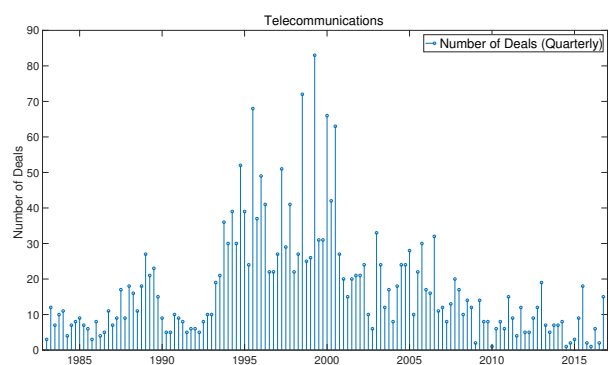
(b) Telecommunications
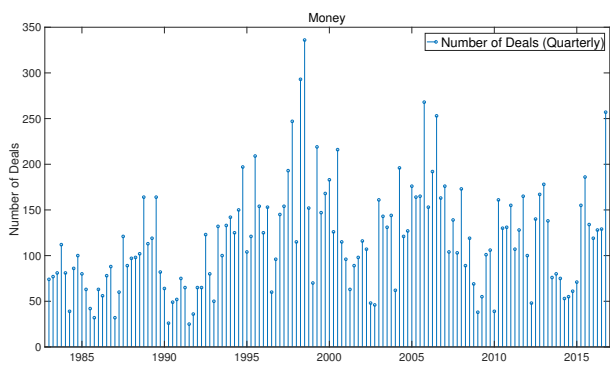
(c) Money
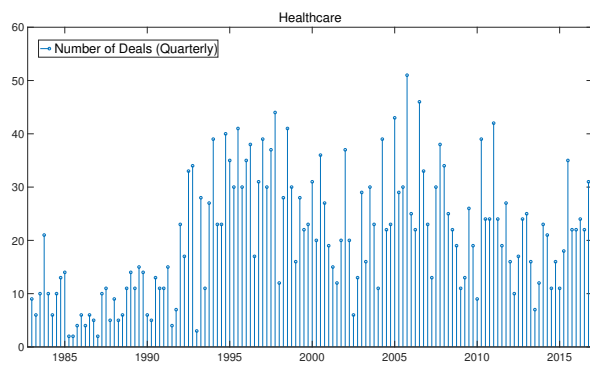
(d) Healthcare

Figure 3: Explanatory Variables and Merger Waves

This figure shows the effect of explanatory variables on the posterior probability of a merger wave as identified by the vector of coefficients $\boldsymbol{\alpha}_{12}^z$. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. The sample period is 1983:Q1 to 2016:Q4, quarterly.
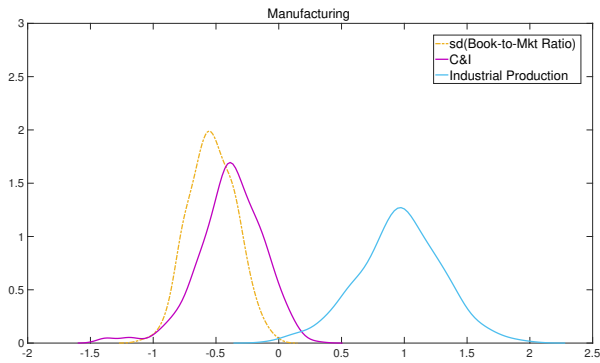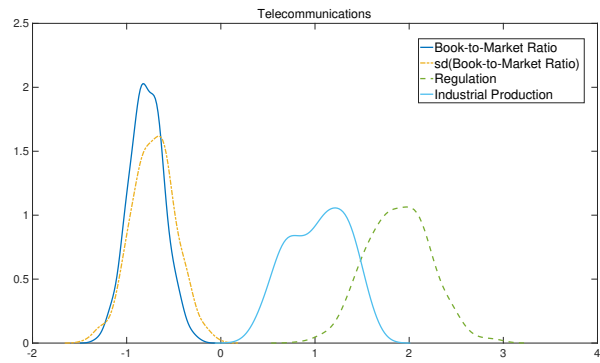
(a) Manufacturing

(b) Telecommunications
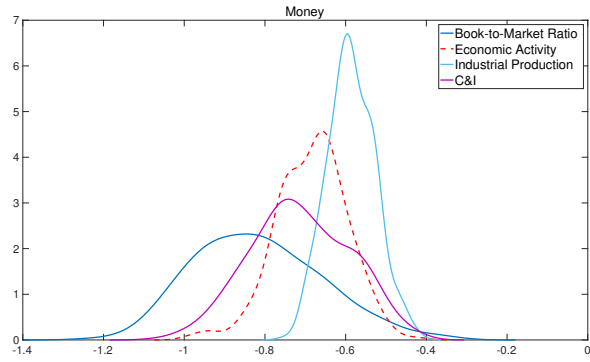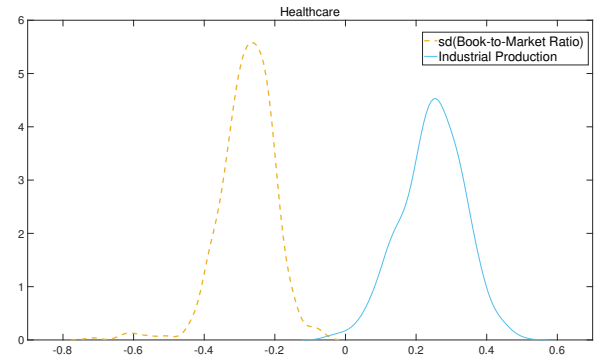
(c) Money

(d) Healthcare

Figure 4: Random Effects Estimates

This figure shows the posterior estimates of the industry-specific random effects for representative sectors, namely Manufacturing, Telecommunications, Healthcare and Money (left panel), and the posterior estimates of the cross-sectional covariance matrix D. The sample period is 1983:Q1 to 2016:Q4.
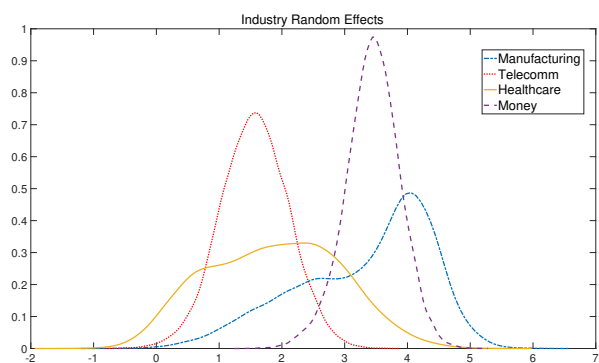
(a) Random-Effects
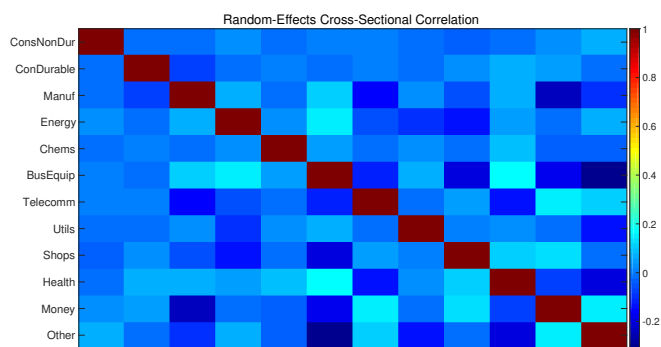
(b) Cross-Sectional Correlation

Figure 5: Posterior Probabilities of Industry Merger Waves

This figure shows the model-implied probability of being in a state of high merger activity. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. The shaded gray areas show the posterior probabilities and the blue lines show the rescaled number of deals for each industry. The sample period is 1983:Q1 to 2016:Q4.
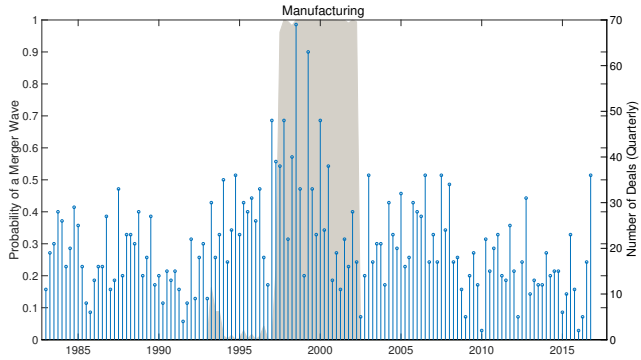
(a) Manufacturing
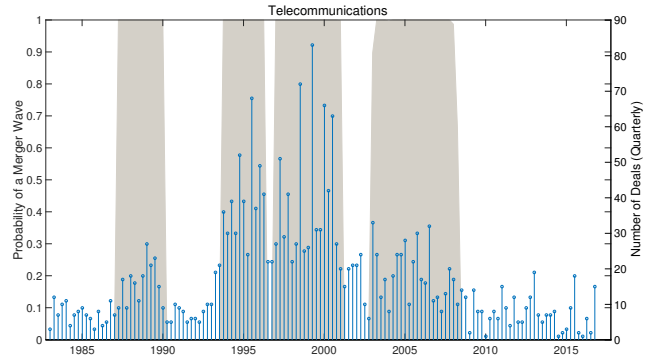
(b) Telecommunications
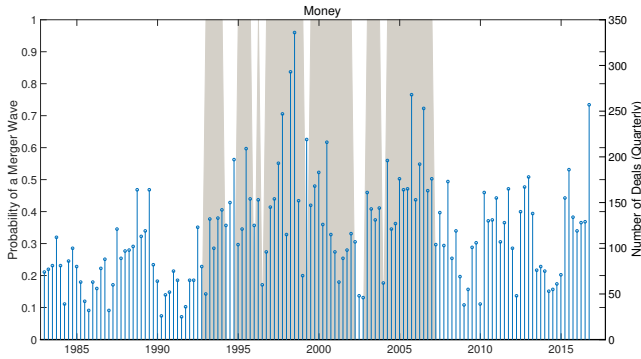
(c) Money

(d) Healthcare

## Figure 6: Explanatory Variables and the Persistence of Merger Waves

This figure shows the effect of explanatory variables on the persistence of merger waves as identified by the vector of coefficients $\boldsymbol{\alpha}_{22}^z$. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. The sample period is 1983:Q1 to 2016:Q4, quarterly.
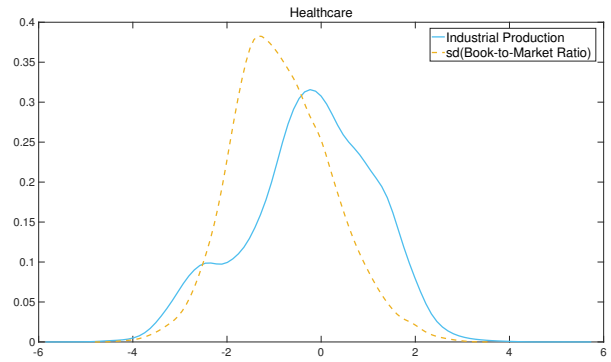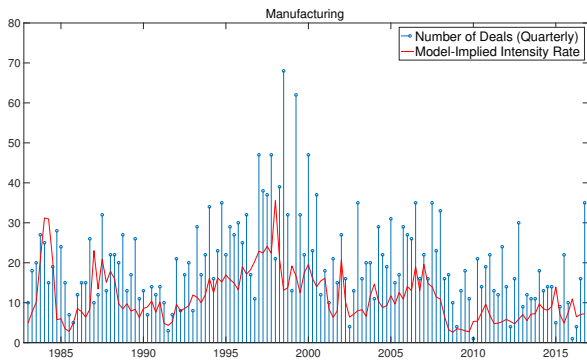

(a) Manufacturing
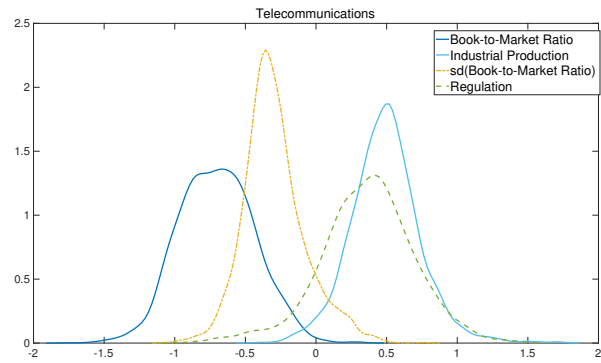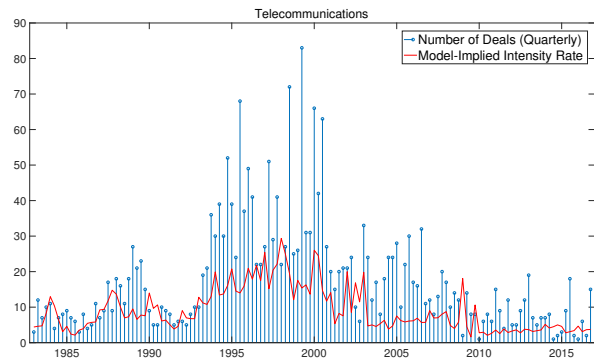

(b) Telecommunications
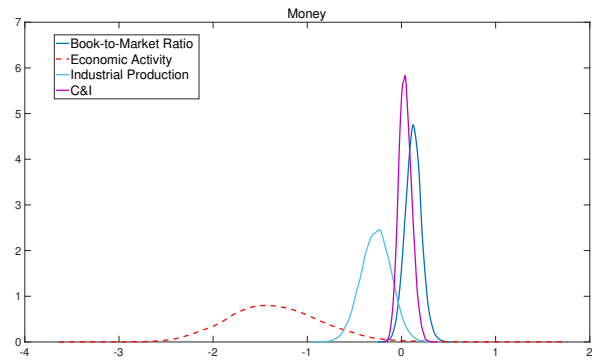
(c) Money

(d) Healthcare

## Figure 7: Merger Activity and Model-Implied Intensity Rates

This figure shows the actual number of M&A deals and the median model-implied intensity rates computed assuming there are two distinct regimes identifying merger waves, i.e. $K = 2$. Deals are measured as the number of all M&A bids announced by US private and public acquirers for US public and private targets in the period 1983:Q1 to 2016:Q4, quarterly. The solid red line is the merger intensity rate implied by the model.
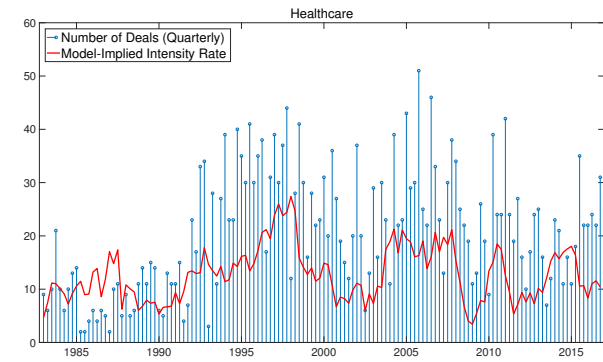

(a) Manufacturing

(b) Telecommunications
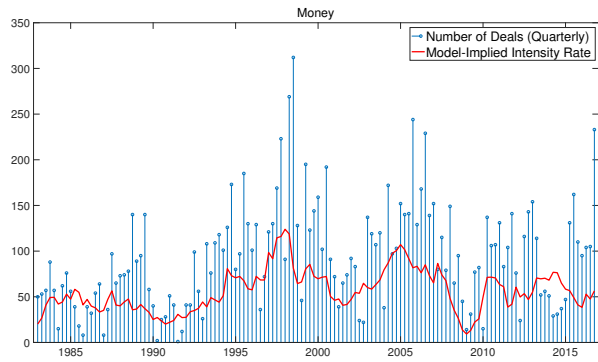
(c) Money

(d) Healthcare

Figure 8: Log Predictive Density Ratio

This figure shows the recursive Log Predictive Density Ratio (LPDR) comparing a model with three regimes against a model with two regimes (top panels), and comparing a model with one regime against a model with two regimes (bottom panels). The left panels show the results for all industries separately and the right panels show the results for the joint predictive density. The sample period is 1983:Q1 to 2016:Q4.
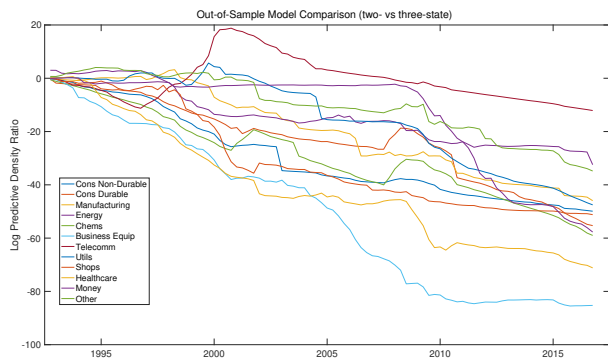
(a) Industry Specific (three vs two states)
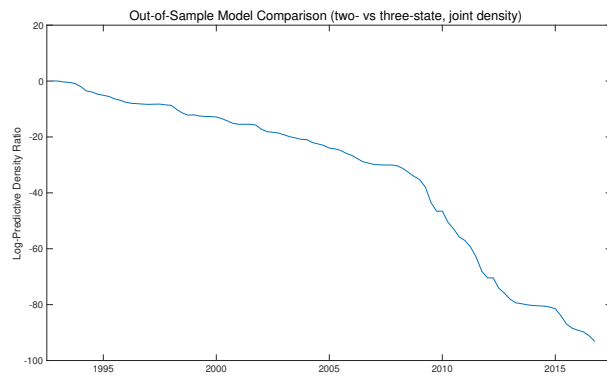
(b) All Industries (three vs two states)
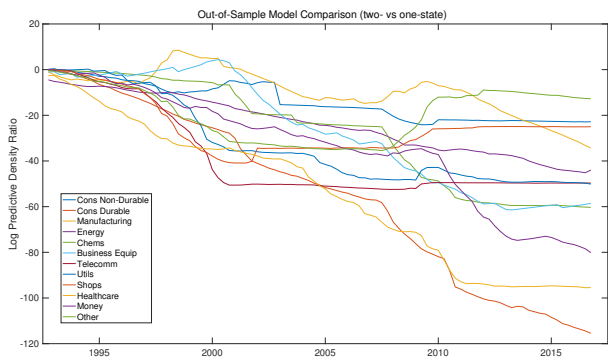
(c) Industry Specific (one vs two states)

(d) All Industries (one vs two states)

Figure A.1: Convergence Diagnostics
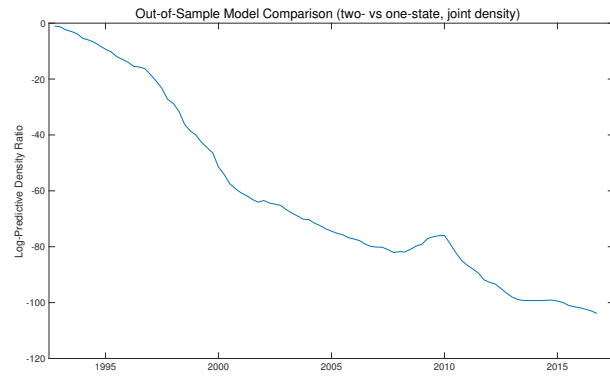
This figure shows some convergence diagnostics for our MCMC estimation scheme. The benchmark specification is with $K = 2$ regimes and the whole set of predictors outlined in Section 3. The left panel shows the recursive mean for the average betas, conditional on both a state of high merger wave (red line) and a state of no-wave (blue line). The running mean is computed by computing the mean for an enlarging window of the MCMC draws of the betas averaged across factors. The right panel shows the autocorrelation function (ACF) of the average betas draws, conditional on both the "no-wave" state (red line) and the "wave" state (blue line). The sample period is 1983:Q1 to 2016:Q4.
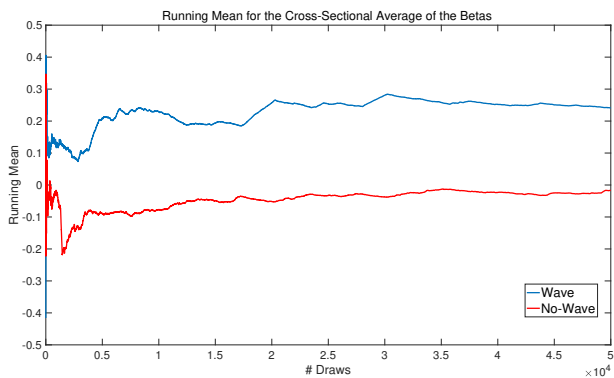
(a) Running Mean
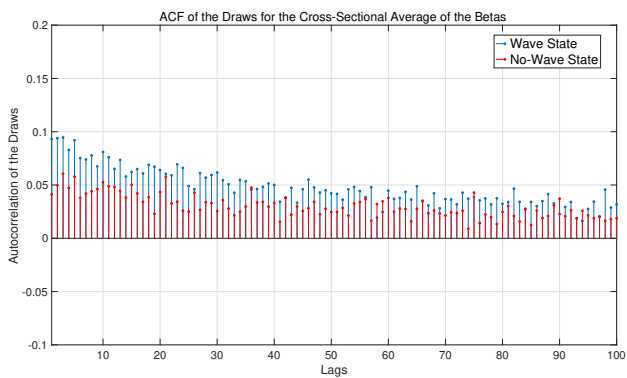
(b) ACF of the Draws

Table 1: Variables Description

This table summarizes the variables we use in our analysis and provides for each variable a description of its measurement, its frequency of observation and the relevant source. Panel A reports the stock valuation variables, related to the behavioral hypothesis. Panel B reports the proxies for industry-specific shocks and capital liquidity, related to the neoclassical hypothesis. Following Harford (2005), these are obtained as industry-specific median absolute annual change, computed across all firms with data available on COMPUSTAT, on the basis of their SIC codes and according to a twelve-industry classification. Panel C reports the aggregate macroeconomic variables.

Panel A: Financial Variables

	Description	Frequency	Source
B/M	Industry-specific book-to-market ratio	quarterly	Kenneth French's data
sd(B/M)	Standard deviation of the industry-specific book-to-market ratio	quarterly	COMPUSTAT
Returns	Industry-specific stock returns (value-weighted)	quarterly	Kenneth French's data

Panel B: Economic Activity

	Description	Frequency	Source
Margin	Margin on sales: Net income/Sales	quarterly	COMPUSTAT
AT	Asset turnover: Sales/Assets	annual	COMPUSTAT
R&D	Research and development expenses scaled by assets at the beginning of the period	quarterly	COMPUSTAT
CAPX	Capital expenditures scaled by assets at the beginning of the period	quarterly	COMPUSTAT
ROA	Return on assets: Net income/Assets at the beginning of the period	quarterly	COMPUSTAT
Sale	Sales	quarterly	COMPUSTAT
C&I	Capital liquidity: spread between the average interest rate on commercial and industrial loans and the Fed Funds rate	quarterly	St. Louis FRED Database
Reg	Dummy variable to indicate whether or not the industry has been subject to a de-regulatory event	quarterly	Viscusi et al. (2005) and Factiva
Margin	Margin on sales: Net income/Sales. We consider the industry specific median absolute change	quarterly	COMPUSTAT

Panel C: Macro-Financial Factors

	Description	Frequency	Source
IP	Industrial production (year-on-year growth)	quarterly	FRED
Credit	Credit spread: the yield spread between 20-year Baa and Aaa corporate bonds	quarterly	FRED
Rf	Real risk-free rate: difference between the one-month T-Bill rate and the CPI inflation rate	quarterly	FRED

Table 2: Deregulatory Events

This table reports major deregulatory initiatives during the sample period 1983:01 to 2016:12. Deregulatory events are considered at the industry level, assigning each event to one of twelve industries based on the SIC code of the bidder at the time of the announcement and according to the twelve-industry classification provided by Kenneth French. The events are constructed from Viscusi et al. (2005) and searching for recent major initiatives in Factiva.

Year	Event	Industry
1984	Cable Television Deregulation Act	Other
	Shipping Act	Other
1987	Elimination of Fairness Doctrine (FCC)	Other
	Sale of Conrail	Other
1989	Natural Gas Wellhead Decontrol Act	Energy
1991	Federal Deposit Insurance Corporation Improvement Act	Money
1992	Cable Television Consumer Protection and Competition Act	Other
	Energy Policy Act	Energy
	FERC Order 636	Utils
1993	Elimination of State regulation of cellular telephone rates	Telecomm
	Negotiated Rates Act	Other
1994	Trucking Industry and Regulatory Reform Act	Other
	Interstate Banking and Branching Efficiency Act	Money
1995	Interstate Commerce Commission Termination Act	Telecomm
1996	Telecommunications Act	Telecomm
	FERC Order 888	Utils
1999	FERC Order 2000	Utils
	Gramm–Leach–Bliley Act	Money

Table A.1: Convergence Diagnostics

This table summarizes the convergence results for the posterior values of the model parameters, estimated over the sample period 1983:Q1 to 2016:Q4. In order to assess inefficiencies, we used the benchmark specification with $K = 2$ regimes and the complete set of explanatory variables outlined in the main text. For each set of parameters, we compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed for the first 20% and the last 50% of the retained MCMC draws after burn-in. The variances of the means are estimated with the Newey and West (1987) estimator using a bandwidth of 4%, 8%, and 15% of the sample sizes, respectively.

Industry	Summary of Convergence Diagnostics			
	i.i.d	Bandwidth of 4%	Bandwidth of 8%	Bandwidth of 15%
Cons. Non-Durb.	0.042	0.349	0.437	0.486
Cons. Durb.	0.893	0.898	0.899	0.905
Manufacturing	0.656	0.866	0.898	0.919
Energy	0.000	0.043	0.114	0.189
Chems	0.249	0.371	0.367	0.342
Business Equip.	0.020	0.056	0.088	0.160
Telecomm	0.000	0.011	0.071	0.124
Utils	0.000	0.031	0.052	0.134
Shops	0.231	0.394	0.474	0.514
Healthcare	0.173	0.416	0.492	0.541
Money	0.132	0.637	0.640	0.641
Other	0.201	0.340	0.550	0.701

References

Amisano, G. and Fagan, G. (2013). Money growth and inflation: A regime switching approach. *Journal of International Money and Finance*, 33:118–145.

Andrade, G., Mitchell, M. L., and Stafford, E. (2001). New evidence and perspectives on mergers. *Harvard Business School Working Paper*.

Barkoulas, J., Baum, C., and Chakraborty, A. (2001). Waves and persistence in merger and acquisition activity. *Economics Letters*, 70:237–243.

Becketti, S. (1986). Corporate mergers and the business cycle. *Economic Review, Federal Reserve Bank of Kansas City*, 71:13–26.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Chib, S. and Winkelmann, R. (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.

Choi, S. and Jeon, B. (2011). The impact of the macroeconomic environment on merger activity: Evidence from us time-series data. *Applied Financial Economics*, 21:233–249.

Coase, R. (1937). The nature of the firm. *Economica*, (4):386–405.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3):299–308.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis*, 51:3509–3528.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modeling. *Biometrika*, 93:827–841.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics 4*.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley, Hoboken.

Geweke, J. and Amisano, G. (2012). Prediction and mispecified models. *American Economic Review*, 102:482–486.

Golbe, D. and White, L. (1993). Catch a wave: The time series behavior of mergers. *The Review of Economics and Statistics*, pages 493–499.

Gort, M. (1969). An economic disturbance theory of mergers. *The Quarterly Journal of Economics*, pages 624–642.

Harford, J. (1999). Corporate cash reserves and acquisitions. *Journal of Finance*, 54:1969–1997.

Harford, J. (2005). What drives merger waves? *Journal of Financial Economics*, 77:529–560.

Holmstrom, B. and Kaplan, S. N. (2001). Corporate governance and merger activity in the us: Making sense of the 1980s and 1990s. Technical report, National Bureau of Economic Research, Working Paper.

Jovanovic, B. and Rousseau, P. L. (2002). The q-theory of mergers. Technical report, National Bureau of Economic Research, Working Paper.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kaufmann, S. (2015). K-state switching models with time-varying transition distributions: does credit growth signal stronger effects of variables on inflation? *Journal of Econometrics*, Forthcoming.

Maksimovic, V. and Phillips, G. (2001). The market for corporate assets: Who engages in mergers and asset sales and are there efficiency gains? *The Journal of Finance*, 56(6):2019–2065.

Maksimovic, V., Phillips, G., and Yang, L. (2013). Private and public merger waves. *The Journal of Finance*, 68(5):2177–2217.

Melicher, R., Ledolter, J., and D'Antonio, L. (1983). A time series analysis of aggregate merger activity. *Review of Economics and Statistics*, 65:423–429.

Mitchell, M. L. and Mulherin, J. H. (1996). The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics*, 41(2):193–229.

Mulherin, J. H. and Boone, A. (2000). Comparing acquisitions and divestitures. *Journal of Corporate Finance*, 6:117–139.

Omori, Y., Chib, S., Shepard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449.

Raftery, A. E. and Lewis, S. M. (1992). Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, 7(4):493–497.

Reid, S. (1968). *Mergers, Managers and the Economy*. McGraw-Hill, New York.

Resende, M. (2008). Mergers and acquisitions waves in the uk: A markov-switching approach. *Applied Financial Economics*, 18(13):1067–1074.

Rhodes-Kropf, M., Robinson, D. T., and Viswanathan, S. (2005). Valuation waves and merger activity: The empirical evidence. *Journal of Financial Economics*, 77(3):561–603.

Rhodes-Kropf, M. and Viswanathan, S. (2004). Market valuation and merger waves. *The Journal of Finance*, 59(6):2685–2718.

Shleifer, A. and Vishny, R. (1992). Liquidation values and debt capacity. *Journal of Finance*, 32:337–347.

Shleifer, A. and Vishny, R. W. (2003). Stock market driven acquisitions. *Journal of Financial Economics*, 70(3):295–311.

Shugart, W. and Tollison, R. (1984). The random character of mergers. *RAND Journal of Economics*, 15:500–509.

Terasvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7(1):119–136.

Town, R. (1992). Merger waves and the structure of merger and acquisition time-series. *Journal of Applied Econometrics*, 7(1):83–100.

Viscusi, W., Harrington, J. W., and Vernon, J. (2005). *Economics of Regulation and Antitrust*. MIT Press, Cambridge, fourth edition.