An Anatomy of Industry Merger Waves*

Daniele Bianchi† Carlo Chiarella‡

First draft: February 2015. This draft: January 19, 2016

Abstract

We propose a novel Markov regime-switching Poisson regression model with endogenous transition distributions to test existing theories on determinants of wave-like patterns in same-industry merger and acquisitions (M&As). We show that the dynamics and persistence of merger waves change substantially in the cross section of deals flow. Valuations and stock returns significantly drive transitions towards periods of abnormally high merger activity for most industries. Except few nuances, industry-specific economic shocks do not sensibly determine waves in market transactions. Also, the empirical analysis demonstrates that deteriorating aggregate economic and credit conditions are expected to be associated with declining M&A activity across industries.

Keywords: M&A, Poisson Regressions, Markov Regime-Switching, Time-Varying Probabilities, Discrete-Choice Models, MCMC

JEL codes: G34, C11, C22, C24, C58

*We are in debt with Sylvia Kaufmann for her help and support. We are also grateful to Yakov Amihud, Armen Arakelyan, Roberto Casarin, Giuseppe Cavaliere, Eric Duca, Maria Teresa González Pérez, Lei Mao, Francesco Ravazzolo, and Linus Siming for their useful comments and suggestions.

†Warwick Business School, University of Warwick, Coventry, UK. Daniele.Bianchi@wbs.ac.uk

‡Colegio Universitario de Estudios Financieros, Madrid, Spain, carlo.chiarella@cunef.edu

1 Introduction

Periodic waves of mergers and acquisitions (M&As) have been an integral part of the capitalist development since its inception, representing an important mechanism for 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.[1]

In this paper, we define a merger wave as a substantial, yet contemporaneous, increase in the intensity of M&A activity at the industry level, and use advanced econometric tools to ask two simple questions that appear to have been neglected so far in the literature. First, we investigate whether the dominant view that merger waves occur independently across industries is backed up by the empirical evidence (see, e.g. Mitchell and Mulherin 1996). Equivalently, we ask whether market transaction data are compatible with the idea that different theory-based factors (e.g. valuation ratios, stock returns and regulatory changes), determine different wave-like patterns in the cross-section of same-industry deals flows. As a result, our first testable hypothesis is whether both industry-specific and aggregate economic conditions affect in the same fashion merger waves across industries. Figure 2 supports our development of formal tests of this hypothesis showing the time series of M&A activity for different industries against a commonly acknowledged

[1]See e.g., Gort (1969), Shugart and Tollison (1984), Golbe and White (1988), Town (1992), Golbe and White (1993), Mitchell and Mulherin (1996), Resende (1996), Harford (1999), Mulherin and Boone (2000), Andrade, Mitchell, and Stafford (2001), Barkoulas, Baum, and Chakraborty (2001), Holmstrom and Kaplan (2001), Maksimovic and Phillips (2001), Jovanovic and Rousseau (2002), Shleifer and Vishny (2003), Rhodes-Kropf, Robinson, and Viswanathan (2005), Rhodes-Kropf and Viswanathan (2004), Floegel, Johanning, and Gebken (2005), Harford (2005), Gugler, Mueller, and Yurtoglu (2006), Resende (2008), Gärtner and Halbheer (2009), Martynova and Renneboog (2009), Choi and Jeon (2011), Gugler, Mueller, and Weichselbaumer (2012), Maksimovic, Phillips, and Yang (2013), and Ahren and Harford (2014).

driving factor such as the cross-sectional standard deviation of firms market-to-book ratios: while the number of deals is clearly related to (the inverse of) variations in valuations (solid blue line) for some industry, including Money, Healthcare, Manufacturing and Consumer Non-Durables, the same mapping can not be sensibly argued for Other, Business Equipment, and Shops industries.[2] Our second question is whether wave-like patterns of merger activity at the industry level are driven by sector-specific economic shocks, rather than being a result of managers market timing, or are ultimately due to aggregate macro-financial conditions. These alternative perspectives show of course an interesting intersection as in this paper we explicitly study whether any differential dynamics in merger waves may derive from a heterogeneous evolution of exposures to standard theory-based determinants.

In methodological terms we make a key choice. Supported by a simple random utility discrete choice model that maps corporate-level M&A decisions to aggregate same-industry merger activity, we propose a novel Markov regime-switching Poisson regression approach with time-varying transition probabilities to investigate the driving factors in the dynamics of deals flows, independently across industries. The main assumption of the model is that the decision to merge depends on parameters that represents the reaction of agents to realizations of an unobservable industry-specific binary state S_t, which identifies mounting evidence of deals, i.e. a merger wave.[3] 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 an abnormally high merger activity generate a higher utility index affecting a firm's 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 do managers value mergers more in some state than others. In that matter, we model merger waves as a first-order Markov state with time-varying

[2]In Figure 2 we report the inverse of the cross-sectional standard deviation of firms market-to-book ratio to better highlight the relationship with M&A activity. The inverse (raw) standard deviation is commonly acknowledged as a measure of confidence (uncertainty).

[3]Notice throughout the paper we used the terms *decision maker*, *firm*, *manager* and *agent* as synonyms.

transition probabilities that explicitly accommodates the endogenous effect of multiple industry-specific and aggregate macro-financial factors. While this modeling framework is fairly general and covers most of the historical examples of merger waves, it also implies that the intensity of merger activity can differ across industries, it is multi-factorial in nature, and unlikely to be entirely captured by any single, observable, industry-specific characteristics or aggregate factor.[4]

Few measurement tools for merger waves have been proposed over the last thirty years (see, e.g. Shugart and Tollison 1984, Town 1992, Golbe and White 1993, Barkoulas et al. 2001, Resende 2008, and Choi and Jeon 2011). Harford (2005) uses a non-parametric approach and identify a wave as a 24-months period with an actual concentration of merger bids in excess of a given threshold. More recently, Maksimovic et al. (2013) identify merger wave years as those in which the percentage of transactions is at least one standard deviation higher than the industry average rate. Alternatively, one may argue that identification of abnormal periods of deals flow can be accomplished through the interaction between time and industry fixed effects within a simple linear regression framework. However, the common feature of these approaches is that determinants z_t of merger waves are investigated separately after wave-like states are identified. As a result, any causality statement is thus to be read as contingent on having full confidence on both waves identification and the fact that determinants can be safely considered as exogenous. However, unless additional assumptions are introduced, such two-step procedures can potentially suffer with endogeneity and misspecification issues that can distort standard hypothesis testing. First, it is likely that agents do not observe the actual state of merger wave, but instead draw inference based on some information set, the content of which is unobservable by the econometrician. The use of a first-step estimate to proxy for this inferential process leads to a regression with measurement error in the explanatory

[4]Following Kaufmann (2015) we use a multi-logit specification for the dynamics of the latent states. This allows to construct an auxiliary state-space representation that allows to separately identify the parameters driving merger waves with respect to the determinants in the observation equation (see Meligkotsidou and Dellaportas 2011 and Holmes and Held 2006 for more details on auxiliary samplers for multi-nomial logistic models).

variables, and thus potential endogeneity. Second, if firms incorporate future expectations of merger waves when determining their investments and firm structure, this will make z_t endogenous. Furthermore, the latent state variable is likely a function of other unobservables, which in turn are correlated with z_t, generating a potential omitted variable bias of unknown direction and magnitude. For example, changes in the production technology for the retail banking sector would likely be correlated with the unobserved state S_t and determinants z_t in the financial industry.[5] Finally, existing approaches potentially suffer with misspecification since mostly tied to linear Gaussian thinking, which makes them of limited value. In fact, M&A activity, as proxied by the number of deals, is a counting process with an unconditional distribution which is discrete and truncated at zero from below.

Differently, we relax the assumption of pure exogeneity in the determinants and propose a parsimonious model that allows for endogeneity in the transition probabilities driving merger waves, also acknowledging the non-Gaussian nature of M&A activity. As far as the model estimation is concerned, we follow 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 both the unknown parameters, the hidden discrete states identifying merger waves, and the time-varying transition probabilities (see Appendix A for more details). By using an MCMC approach, we are able to make inference on parameters and latent states in a single step, developing a robust finite-sample testing framework that helps to generate posterior distributions of virtually any function of the model outputs. More prominently, posterior estimates allow to test hypothesis on the nature, dynamics, and structure of industry merger waves in a unified setting, something that earlier literature did not provide.

[5] One may argue some standard driving factor such as a regulatory change can be unequivocally seen as an exogenous event. However, in principle, also deregulation can be interpreted as an endogenous determinant of merger waves on itself. In fact, regulatory changes are typically preceded by poor industry performance and therefore can be partly predicted by performance variables.

Empirically, we first test two common competing explanations of merger waves, i.e. the neoclassical and the behavioral, by focusing 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 2014. 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. The neoclassical explanation of merger waves suggests M&A activity primarily occurs in response to technological, regulatory and/or economic shocks that trigger restructuring and consolidation at the industry level (see, e.g. Coase 1937, Gort 1969, Shleifer and Vishny 1992, Harford 1999, Andrade et al. 2001, and Holmstrom and Kaplan 2001, Maksimovic and Phillips 2001, Jovanovic and Rousseau 2002, and Harford 2005). Instead, the behavioral hypothesis builds on the observed positive correlation between stock valuations and merger activity and suggests that M&As are driven by firm-specific and market-wide valuations, or managerial market timing. Different explanations lie either on irrationality of the stock market and self-interest of management, or on correlated valuation errors in rational markets (see e.g., Shleifer and Vishny 2003, Rhodes-Kropf and Viswanathan 2004, Rhodes-Kropf et al. 2005 and Maksimovic et al. 2013). In addition to these mainstream theories we also investigate the explanatory power of aggregate macro-financial conditions (see, e.g., Melicher, Ledolter, and D'Antonio 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).

We report some novel findings. First, we show that the dynamics and persistence of merger waves change substantially across industries over the last thirty years. While a state of abnormally high merger activity lasts few years in industries such as Shops, Other and Money, the same lasts more than a decade for the Telecomm, Healthcare and Business Equipment industries. A simple cross-industry correlation analysis shows that merger waves are not synchronous among industries. This implies that waves-like patterns are industry specific, and any generalization at the aggregate market level can potentially make inferential statements misleading.

Second, we show that changes in valuations and stock returns sensibly drive merger waves. While positive for most industries, market-to-book ratios for Shops, Money and Other have a negative effect on merger waves, consistent with the idea that in these industries M&As are mostly counter-cyclical and occurs for restructuring purposes. This is confirmed by the opposite effect of past returns variance on the transition probabilities. Except few nuances, e.g. returns on assets, economic shocks do not represent strong determinants of abnormal periods of merger activity. As a result, the neoclassical hypothesis on the origin of merger waves can not be significantly supported in our sample. Interestingly, a recession dummy and the spread between 20-years Baa and Aaa-rated corporate bonds, show that deteriorating aggregate economic and credit conditions are expected to be associated with declining intensity of M&A deals. This argues for a more general "merger activity-economic prosperity" theory (see, e.g. Reid 1968).

Third, we provide evidence that regressors related to the behavioral theory of merger waves have a higher in-sample explanatory power for the time-series of deals flow. More specifically, an in-sample model comparison exercise based on log marginal likelihoods favor a model with the cross-sectional average and variation of firm-level market-to-book ratios, stock returns and aggregate dividend-yield as explanatory variables. Finally, an out-of-sample forecasting exercise, based on both a standard Root Mean Squared Error (RMSE) measure and more precise log predictive scores, shows that the same set of behavioral-related regressors significantly helps predict M&A activity up to several months ahead for the majority of industries, at least in the context of our Markov regime-switching Poisson regression model.

The paper is structured as follows. Section 2 lays out our modeling framework. Section 3 describes the estimation strategy. Next, Section 4 discusses the data used. Section 5 represents the heart of the paper and contains our findings on heterogeneous merger waves estimates and determinants, with special emphasis on the dichotomy neoclassical vs. behavioral hypothesis, independently across industries. Section 6 concludes.

2 Modeling Same-Industry M&A Activity

In this section we start with a simple, textbook, random utility discrete choice model that allows to map corporate-level M&A decisions to aggregate same-industry merger activity. Thus, we propose a novel Markov regime-switching Poisson regression approach with endogenous transition distributions to test existing theories on merger waves.

Let $i = 1, ..., N$ be the number of firms operating in one of the $m = 1, ..., M$ industries. Without lost of generality we can think of a decision to engage in M&A as the outcome of a binary discrete choice $j \in [0, 1]$. Each bidder decides to engage in a deal at time t on the basis of an information set which contains lagged values of industry-specific information w^m_{t-1}, aggregate variables x_{t-1}, unobservable individual attributes $\epsilon^m_{ij,t}$, and parameters that represent the reaction of agents to realizations of an unobservable industry-specific binary state S_t, which identifies a merger wave. The utility of firm i from the choice j at time t can now be defined as;

$$U^m_{ij,t} = z^m_{t-1}\beta^m_{S_t} + \epsilon_{ij,t}, \qquad i = 1, ..., N, \qquad j = 0, 1, \qquad m = 1, ..., M \tag{1}$$

with $z^m_{t-1} = \left(x_{t-1}, w^m_{t-1}\right)$, $\beta^m_{S_t}$ the state-dependent sensitivities to the regressors z^m_{t-1}, and $z^m_{t-1}\beta^m_{S_t}$ representing the "systematic" component of the utility index. As typical in random utility models, we assume the reaction to industry-specific or aggregate determinants $\beta^m_{S_t}$ is constant across firms. This implies that all of the individual factors, such as for instance the risk aversion of the management, are incorporated in the unobservable attributes $\epsilon_{ij,t}$ (see, e.g. Guimarães, Octavio, and Woodward 2003, Dube, Hitsch, and Rossi 2009, and Burda, Harding, and Hausman 2011, for related examples).

The dynamics of the latent state is endogenously determined by current values of the determinants $z^m_t = (x_t, w^m_t)$. We assume $S_t = k$, $k = 1, \cdots, K$, $t = 1, \cdots, T$ follows a first-order Markov process with time-varying transition probabilities specified through a multi-logit function.[6] We decompose the effect of determinants in two components. A time varying component $z_t^m \alpha^z_{St-1,S_t}$ capturing the pure effect of the determinants on switching probabilities, and a time-invariant component α_{S_{t-1},S_t} capturing the state persistence;

$$p\left(S_t = k | S_{t-1} = l, z_t^m, \alpha\right) = \xi_{lk,t} = \frac{\exp\left(z_t^m \alpha^z_{lk} + \alpha_{lk}\right)}{\sum_{j=1}^{K} \exp\left(z_t^m \alpha^z_{lj} + \alpha_{lj}\right)}, \tag{2}$$

with $\alpha' = \left(\alpha^{z,\prime}_{S_{t-1},S_t}, \alpha'_{S_{t-1},S_t}\right)$ an $(N+K)-$dimensional vector of parameters. The hidden nature of S_t prevent the decision maker to map deterministically z_t^m into an abnormally high state of merger activity. Although our model can be generalized to have multiple regimes, throughout the paper we assume there are two states, i.e. $S_t = 1, 2$. This assumption is backed by a formal marginal likelihood test as shown in Section B of the Appendix. The reference state is defined as the state in which the number of deals cannot be classified as falling within a merger wave. For identification purposes, the parameters governing the transition to the state of no merger wave are restricted to be equal to zero, i.e. $\left(\alpha^{z,\prime}_{S_{t-1},1}, \alpha'_{S_{t-1},1}\right) = (0', 0')$.[7]

The specification (2) is rather general as it involves standard implementations as special cases. By restricting the effect of the covariate z_t^m to be independent on past states, i.e. $\alpha^z_{lk} = \alpha^z_k$, we assume the dependence of the state of merger waves is only governed by the time-invariant component α_{lk} (see, e.g. Filardo 1994 and Amisano and

[6]Alternative approaches to model time-varying transition probabilities are based on a probit specification (see, e.g. Filardo 1994, Filardo and Gordon 1998, and Amisano and Fagan 2013. However, a probit specification implies that the transition probabilities should be numerically evaluated at each time t if the latent state indicator is not known, as in our model. For a time series of length T, this amounts to evaluate $TK(K-1)K-1$-dimensional multivariate normal integrals. Although numerically feasible, this is computational intensive and massively counterbalance the linearity inherited by the latent auxiliary state-specification normally exploited to estimate a probit regression.

[7]For a model with two states this yields;

$$p\left(S_t = 1 | S_{t-1} = l, z_t^m, \alpha\right) = \xi_{l1,t} = \frac{1}{1 + \exp\left(z_t^m \alpha^z_{l2} + \alpha_{l2}\right)}, \qquad l = 1, 2 \tag{3}$$

As such, the vector of parameters $(\alpha^z_{l2}, \alpha_{l2})$ is the only set of parameters we need to estimate to characterize the transition dynamics of merger waves.

Fagan 2013). Further, a standard time-homogeneous Markov regime-switching model can be recovered as a special case by restricting $\alpha^z_{lk} = 0, \forall l, k$. If, instead, we restrict both $\alpha^z_{lk} = \alpha^z_k$ and $\alpha_{lk} = \alpha_{lk}$, 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 z^m_t, and the transition mechanism is independent on the lagged prevailing state. Figure 1 depicts the dependence structure underlying the model dynamics.

[Insert Figure 1 about here]

The decision to embark in a deal for a given industry is driven by past financial and/or economic factors and the agent's reaction to realizations of the unobservable state of merger wave β_{S_t}. Conditional on S_t, the index utility $U^m_{ij,t}$ is not affected by current values of theory-based determinants. However, z^m_t has a direct effect on the one-step ahead utility $U^m_{ij,t+1}$, and thus is endogenous. This captures the idea that firms incorporate expectations of current merger waves in determining future investment and capital structure decisions, which in turn affects the merger activity within industries through the firm-level propensity to merge. For identification purposes we assume the decision of not to engage in a deal $j = 0$ is the benchmark, i.e. $U^m_{i0,t} = \epsilon^m_{i0,t}$, such that

$$M^m_{i,t} = \begin{cases} 1 & \text{if} \quad z^m_{t-1}\beta^m_{S_t} + \epsilon_{i1,t} \geq \epsilon_{i0,t} \\ 0 & \text{if} \quad z^m_{t-1}\beta^m_{S_t} + \epsilon_{i1,t} < \epsilon_{i0,t} \end{cases} \tag{4}$$

To reflect that (4) uncovers the choice with the highest expected utility, $\epsilon_{ij,t}$ follows a type I extreme value distribution where the location and scale parameters are normalized to 0 and 1, respectively.[8] Based on McFadden (1974) we can show that if the *ith* bidder is utility maximizer, his probability of engaging in M&A at time t, conditional on the

[8]Assuming the firm-specific idiosyncratic term follows an i.i.d. type I extreme value distribution means that the probability density function of $\epsilon_{ij,t}$ is defined as

$$p(\epsilon) = \exp(-\epsilon)\exp(-\exp(-\epsilon))$$

which is also known as standard Gumbel distribution.

state S_t and the information at time t is given by

$$\pi_{S_t,t}^m = Prob\left(M_{i,t}^m = 1\right) = \frac{\exp\left(z_{t-1}^m \beta_{S_t}^m\right)}{1+\exp\left(z_{t-1}^m \beta_{S_t}^m\right)}, \tag{5}$$

This represents a familiar Conditional Logit Model (CLM) formulation. Let us denote with y_t^m the number of deals, $y_t^m = \sum_{i=1}^N M_{i,t}^m$ at time t for the mth sector. Now we can estimate the parameters of (5) by maximizing the following log-likelihood

$$\log L\left(\mathbf{y}_{1:T}^m | \mathbf{S}_{1:T}, \mathbf{z}_{1:T}^m, \beta^m\right) = \sum_{t=2}^T y_t^m \log \pi_{S_t,t}^m, \qquad m = 1, ..., M, \tag{6}$$

with $\mathbf{y}_{1:T}^m = (y_1^m, \cdots, y_T^m)$, $\mathbf{z}_{1:T}^m = (z_1, \cdots, z_T)$ and $\mathbf{S}_{1:T} = (S_1, \cdots, S_T)$ the time series of deals and binary states. As shown in Guimarães et al. (2003), such likelihood function is equivalent to that of a Poisson model which takes as a dependent variable the number of merger deals y_t^m, and the intensity rate is a direct function of the systematic component of the utility index ;

$$y_t^m | \lambda_{S_t,t}^m \sim Poisson\left(\lambda_{S_t,t}^m\right), \qquad \lambda_{S_t,t}^m = \exp\left(z_{t-1}^m \beta_{S_t}^m\right), \tag{7}$$

with the dynamics of S_t specified by (2). We use the Markov regime-switching Poisson regression with time-varying transition probability (7) as an approximation of the aggregation of the individual decision problem (1)-(4) and tests existing theories on merger waves based on the significance of the estimated parameters α_{12}^z for different set of theoretically justified covariates.

3 Estimation Strategy

Estimation of Markov regime-switching Poisson regressions with endogenous transition distributions is a rather challenging problem. Unlike for linear regression frameworks, maximum likelihood is not available in closed form when the latent state S_t is integrated out (see, e.g. Kim, Piger, and Startz 2008 for a linear regression example). Each evaluation of the likelihood function requires some numerical method for solving the necessary high-dimensional integration. In this paper, we follow 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 both the unknown parameters, the latent state identifying a merger wave, and the time-varying transition probabilities (see Appendix A for more details). For the ease of exposition, in the following we drop the superscript indicating the mth industry. The regression parameters for the intensity rate are gathered into the vector $\beta = (\beta_1, \ldots, \beta_K)$, where $\beta_k = (\beta_{1k}, \ldots, \beta_{pk})$, for $k = 1, \ldots, K$. The parameters governing the transition probabilities are denoted by $\alpha = \{\alpha_k | k \in \mathcal{K}_{-k_0}\}$ with k_0 the reference state, and $\alpha_k = (\alpha^z_{1k}, \ldots, \alpha^z_{Kk}, \alpha_{1k}, \ldots, \alpha_{Kk})$.

3.1 Prior Specification

For the Bayesian inference to work, we need to specify the prior distributions for the model parameters. For a given state $S_t = k$, with $k = 1, \ldots, K$, the prior structure is conjugate as the function for the (log of) intensity rate reduces to standard multiple linear regressions. The prior for the betas is independent on any assumption about the transition probabilities of the states. For each state $k = 1, \cdots, K$ we assume a normal conjugate prior;

$$p(\beta) = \prod_{k=1}^{K} p(\beta_k) = \prod_{k=1}^{K} \mathcal{N}_p(b_k, B_k) \tag{8}$$

with b_k and B_k the location and scale hyper-parameters, respectively. The prior specification (8) assumes the state-dependent regression parameters for the intensity rates to be independent to each other. This makes the prior to be invariant to state permutations. As far as the transition probabilities are concerned, the multi-logit specification for the transition probabilities (2)-(3) allows to assume a Gaussian prior distribution for the parameter vector α_k;

$$p\left(\alpha\right) = \prod_{k \in \mathcal{K}_{-k_0}} p\left(\alpha_k\right) = \prod_{k \in \mathcal{K}_{-k_0}} \mathcal{N}_N\left(a_k, A_k\right) \tag{9}$$

in which, again, the independence across states makes the prior invariant to state permutations. The independence across states allows to specify prior hyper-parameters independently across regimes, such that one prior can be imposed to be more informative for a state of, say, merger wave. In the following we take an agnostic approach and impose the same degree of non-informativeness across states.

3.2 Posterior Simulation

In order to estimate the parameters and the latent states we implement a data augmentation scheme (see Tanner and Wong 1987 and Frühwirth-Schnatter 1994) which relies on a complete likelihood function, that is the product of the data and state variable densities, given the parameters. Let us denote with $\mathbf{y}_{s:t} = (\mathbf{y}_s, \ldots, \mathbf{y}_t)$, $s \leq t$, a collection of vectors $\mathbf{y}_u$. The collections of parameters are defined as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$, respectively, where $\theta_k = (\beta_k, \alpha_k)$, $k = 1, \ldots, K$, are the state-specific parameters. Given the model structure (7), the completed data likelihood is

$$p\left(\mathbf{y}_{1:T}, \mathbf{S}_{1:T} | \mathbf{z}_{1:T}, \boldsymbol{\theta}\right) = \prod_{k,l=1}^{K} \left(\prod_{t=1}^{T} y_t!\right)^{-1} \exp\left(-\sum_{t=1}^{T} \exp\left(\beta_t' z_{t-1}\right) + \sum_{t=1}^{T} y_t \beta_t' z_{t-1}\right) \xi_{lk,t}^{N_{lk,t}} \tag{10}$$

with $N_{lk,t} = \mathbb{I}_{\{l\}}(S_{t-1})\,\mathbb{I}_{\{k\}}(S_t)$ and $\beta_t = \sum_{k=1}^{K} \beta_k \mathbb{I}_{\{k\}}(S_t)$. Combining the prior specifications (8)-(9) with the complete likelihood (10), we obtain the posterior density

$$p(\boldsymbol{\theta}, \mathbf{S}_{1:T}|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}) \propto p(\mathbf{y}_{1:T}, \mathbf{S}_{1:T}|\boldsymbol{\theta}, \mathbf{z}_{1:T})\, p(\boldsymbol{\theta}) \tag{11}$$

Since the joint posterior distribution is not tractable analytically, the estimator of the parameters cannot be obtained in closed form. The random draws from the joint posterior distributions are obtained through a Gibbs sampler algorithm (Geman and Geman 1984). Following Frühwirth-Schnatter and Wagner (2006) and Kaufmann (2015) we design an approximate sampling scheme which helps to get rid of both non-normality and non-linearity in an efficient way. A detailed description of the Gibbs sampler and its convergence properties are provided in Appendix A and C, respectively. Posterior estimates are based on 50,000 draws with a burn-in of 10,000 and thin value of 5. Such a choice of the number of draws keeps the computational burden relatively low, at the benefit of inference precision as shown in Table C.1 and Table C.2 in Appendix C.

4 Data and Descriptive Statistics

The sample contains all announced bids for US private and public acquires that were announced in the period from 1983 to 2014, for which the bidder did not previously own a majority interest in the target and is seeking to obtain a majority interest through the transaction. Data on M&A deals flow are collected from Thomson One Banker and complemented with firm-level stock market and accounting data from the Center for Research in Security Prices (CRSP) and COMPUSTAT databases, respectively. Macroeconomic factors are collected from the Federal Reserve Economic Data series or 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, or an acquisition of partial or remaining interest. The sample includes overall 60,305 observations for which deal value is disclosed.

M&A deals are aggregated at the industry level monthly, 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 month and for each of the remaining nine industries, with a set of variables proxying for economic shocks and market valuations.

Industry-specific economic shocks are associated to the neoclassical hypothesis on merger waves (e.g. Coase 1937, Gort 1969, Shleifer and Vishny 1992, Harford 1999, Andrade et al. 2001, and Holmstrom and Kaplan 2001, Maksimovic and Phillips 2001, Jovanovic and Rousseau 2002, and Harford 2005). Following Harford (2005) we consider: 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. CAPX, 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). In order to detect economic shocks, for all these variables we consider the industry specific median absolute annual change, computed across all firms with data available on COMPUSTAT in each industry. Calculating these variables at the industry level we allow economic shocks to have different effects and magnitudes across industries. In addition, other factors related to the neoclassical hypothesis, are the contemporaneous capital liquidity (CL), 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 structure consists also of variables that capture market timing and valuations which are typically identified by the behavioral hypothesis as driving factors for merger activity (see e.g., Shleifer and Vishny 2003, Rhodes-Kropf and Viswanathan 2004, and Rhodes-Kropf et al. 2005). The set of factors includes the industry-specific average market-to-book ratio (MB), its cross sectional standard deviation (std(MB), computed across all firms with data available on COMPUSTAT for each industry), the value-weighted industry stock returns (Returns), as well as the realized variance of aggregate market returns (svar) and the aggregate dividend yield (dy), computed across all firms with data available on COMPUSTAT.

In addition to these mainstream determinants we also consider a set of variables that capture aggregate macroeconomic conditions (see, e.g., 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: a dummy variable that indicates NBER recession periods from peak through the trough, industrial production growth as a measure of changes in aggregate output, the term spread (measured as the yield spread between the 10-year government bond and 3-month T-bill), the credit spread (measured as the yield spread between Moody‘s 20-years Baa and Aaa-rated corporate bonds), the real risk free rate (proxied by the difference between the 1-month T-bill rate and the CPI log inflation rate), and the aggregate market-to-book ratio. The fact that we use the aggregate market-to-book ratio as macroeconomic variable partly overlaps with the behavioral hypothesis. 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 the monthly frequency as a trade-off between having enough granularity in the data and keeping enough information, namely a sufficient number of deals for each data point and for each industry. In this respect, the annual frequency of the balance-sheet items complicates the empirical analysis as we need to investigate the determinants of merger activity on a monthly basis. We interpolate through a cubic spline the values for all dates over the period, using end of year values for accounting items.[9]

[9] In order to ensure an coherent comparison across theories we collected the data on both financial

The interpolation method has the advantage of producing a smooth implied decision process, which is more consistent with our random utility model, and, in particular, avoids jumps in the implied intensity rate resulting from impounding the entire change in merger activity to one day at the end of each period.

[Insert Table 2 about here]

Table 2 provides some insights on the composition of our sample. Panel A reports summary statistics on the distribution of the monthly deals flow in each industry. Pressure to embark in M&A activity according to the neoclassical and behavioral hypotheses are considerably different across industries. For instance, the financial sector shows almost ten times the intensity rate of M&A decisions with respect to Consumers (both Durables and Non-Durables), Chemicals and Utilities. Business Equipment ranks second in terms of average deals flow. Interestingly, those sectors with the highest average activity, are those showing the highest unconditional volatility. The persistence of the merger activity is rather low, except for Business Equipment, Money and Other. The auto-regressive coefficient of an AR(1) model for the demeaned series gets nowhere higher than 0.6 for more than a half of the sectors. Sample average values for the industry-specific economic shocks are reported in Panel B. Except few nuances, average values are heterogeneous across industries. For instance, shocks on ROA for the Healthcare industry are ten times higher than for the financial sector. R&D and CAPX have been relatively stable in the cross-section, on average. High cross-sectional variation is also shown by both AT and Emp, with the Energy and the Healthcare industries having the highest deviations from median values, on average.

Finally, Panel C of Table 2 shows the sample averages of market valuation variables for each industry. Coupled with the average number of deals in Panel A, these sample

variables and economic shocks at the same frequency, i.e. quarterly, then we interpolate both set of determinants to approximate monthly time series.

averages make clear that there is no direct, unconditional, mapping between market-to-book ratios and average merger activity. In fact, the opposite happens for Utilities, in which despite showing a rather high sample valuation, the average number of deals is lower than Shops, which in turns are actually undervalued. Also, the cross-sectional within-industry value-weighted variation of the market-to-book ratio does not show any interesting heterogeneity. As a whole, in unconditional terms, there is no clear mapping between theory-based justifications of merger activity and the actual sample average of deals flow.

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 de-regulatory events.

[Insert Table 3 about here]

Table 3 documents major de-regulatory initiatives during the sample period 1983:01-2014:12. De-regulatory 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, Harrington, and Vernon (2005) and searching for recent major initiatives in Factiva. A comparison with time series of merger activity may suggest a relationship which will be formally explored further in the next section.

5 Empirical Results

This section represents the heart of the paper and reports our findings on merger waves estimates and determinants. We exclude from the final sample the consumer durables, chemicals and utilities sectors as the monthly average number of deals in these sectors is not sufficient to ensure a clear regimes identification. Figure 2 shows that the time series of market transaction data (grey bars) is not synchronous across industries but varies heavily in the cross-section. For instance, while the deals flow for the consumer non-durables industry peaks at $y_t^{NoDur} = 15$ on late 90s, the intensity of M&A is equal to 150 in the same period for Money. Also, different industries show different persistence in periods of high merger activity. For instance, the deals flow is remarkably high for almost ten years for the telecommunications industry, while lasts few years for Money and Other. As a result, it is likely that aggregating information at the market level "dilutes" the effect of industry-specific determinants on merger waves.

We first test the significance of determinants of merger waves with particular emphasis on the dichotomy neoclassical vs. behavioral hypothesis vs. macroeconomic factors, for each industry independently, and show the corresponding model-implied state probabilities. Then we show the time series of merger intensity rates against the data and formally test through marginal likelihood evidence the goodness-of-fit obtained by using the alternative theory-based sets of regressors. Finally, we show the predictive performance of the model for each industries and across different forecasting horizons. We assume there two states $S_t = 1, 2$. This assumption is backed by a formal marginal likelihood test as shown in Section B of the Appendix.

5.1 Determinants of Merger Waves

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 the $N-$dimensional vector $\alpha_{12}^z = \left(\alpha_{12,1}^z, \cdots, \alpha_{12,N}^z\right)$. For instance, a positive and significant $\alpha_{12,1}^z$ implies that the first factor can explain a regime switch towards a merger wave. On the other hand, if $\alpha_{12,1}^z$ is not statistically different from zero, it means that the first factor does not sensibly affect regimes of merger activity. Figure 3 shows the results for the behavioral theory.

[Insert Figure 3 about here]

Except for Consumer Non-Durables and Manufacturing the market-to-book ratio is a sensibly driver of merger waves, consistent with the evidence provided by Shleifer and Vishny (2003), Rhodes-Kropf and Viswanathan (2004), and Rhodes-Kropf et al. (2005). Interestingly, for Shops, Money and Other, valuation ratios have a negative effect on merger waves. This is coherent with the idea waves in M&A can also be counter-cyclical and occurs for restructuring purposes in some industry, as confirmed by the opposite effect of the past variance of returns on the transition probabilities for the same industries. Past realized returns are significant and positive determinants of merger waves for the Energy, Healthcare and Business Equipment industries. Differently, dividend-yield does not represent a significant driving factor. Figure 4 shows the results of testing the neoclassical hypothesis on merger waves. With only few exceptions, e.g. AT and ROA, there is no clear effect of economic shocks on the transition to a merger wave. The returns on assets are negative related with abnormal M&A activity in the Energy industry, while ROA is positively related to the state indicator in the Telecomm, Healthcare and Manufacturing industries.

[Insert Figure 4 about here]

Overall, although relevant for some instances, shocks to industry-specific economic conditions cannot be sensibly identified as driving factors of periods of merger waves. As a result, our empirical evidence does not support the neoclassical theory on merger waves for most industries. Finally, Figure 5 shows the effect of aggregate macro-financial variables on merger waves. As far as the business cycle is concerned, while neutral for most two-third of the industries, deteriorating aggregate economic conditions have a negative effect on the propensity to engage in a deal within the Business Equipment, Consumer Non-Durables and Money industries. For Money and Consumer Non-durables the negative effect of business cycles is confirmed by the positive influence on the propensity to merger of output growth, as proxied by changes in industrial production. The latter also shows a positive effect on abnormal periods of M&A activity for Other, although recessions do not seem to play a role in that matter. Similarly, aggregate credit conditions, as proxied by the credit risk premium, have a negative effect for all but Business Equipment and Consumer Non-Durables. This is partly consistent with the idea that credit conditions have a significant effect on the agents propensity to embark in M&A decisions.

[Insert Figure 5 about here]

Although other macro-financial factors such as the Term spread and aggregate market valuations, as proxied by the aggregate market-to-book ratio, do not play a significant role in the dynamics of merger waves, the fact that aggregate economic and credit conditions are significant driving factors is consistent with a more general "merger activity-economic prosperity" theory to complement the behavioral explanation (see, e.g. 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).

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 in a standard cross-sectional regression would likely average out such longitudinal variation. We now investigate the model-implied probabilities of a merger wave conditional on different sets of regressors and for each industry. As far as the determinants of merger waves are concerned, industry-specific financial conditions seem to be more plausible as driving factors of abnormal merger activity. Figure 6 shows the posterior median probabilities of being in a state of merger wave across industries for a model that ground on the behavioral hypothesis for merger waves. The initial wave periods are mostly located across the second half of the 90s and the beginning of 2000s. This is consistent with the results in

Harford (2005). However, two important unexplored features emerge.

[Insert Figure 6 about here]

First, coherent with the time series behavior of deals flow, the length of model-implied merger waves is not homogeneous across industries, as for some industries (e.g. Energy, Telecommunications, Healthcare and Business Equipment) waves are much more persistent than for Shops, Money, Other, and Consumer Non-Durables. Such heterogeneity in persistence would be completely ignored by pooling together information on deals flow and investigate 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 Consumer Non-Durables mostly coincided with the presidency of Ronald Reagan, and the economic prosperity of the mid- to late-80s and the second half of the 90s, the same explanation would not apply for the Other industry. Differently, Telecomm would be consistent with the idea that merger waves followed the economic recession of 1990-91, as confirmed by a positive effect of the recession indicator on the propensity to engage in a deal (see, Figure 5). Figure 7 reports the posterior median probabilities of a merger wave one would obtain if one believe in the neoclassical hypothesis, in the context of our model.

[Insert Figure 7 about here]

Except for Manufacturing and Healthcare industries, the overlapping with the merger waves probabilities implied by the behavioral-theory is minimal. Spikes in deals flow are almost entirely missed for Money, and partly missed for Consumer Non-Durables and Telecomm. As far as Other is concerned, while the model-implied probability of a merger wave is close to one towards the end of the sample, the corresponding actual number of deals is around its unconditional mean. A similar inconsistency appears for the

Energy industry, where a neoclassical-related model specification identifies abnormally high merger activity during the second half of the 80s, despite an actual number of deals around its sample minimum. Finally, Figure 8 shows the posterior median probabilities of being in a state of merger wave across industries for a model that depends on aggregate macro-financial factors.

[Insert Figure 7 about here]

Interestingly, except for Money, Consumer Non-Durables, Energy and Shops the identification is consistent with the behavioral theory. For instance, the main wave that characterized the Telecomm industry from late 90s to early 2000s is captured under both specifications. Similarly to the behavioral-related specification, the persistent period of abnormally high merger activity within the Business Equipment and Healthcare industries is captured closely by the model estimates.

Notably, there are few "short-lived" periods identified by the model as a merger wave. This is due to the fact that the grey areas report the "filtered" probabilities which are computed conditioning on the information available on deals and determinants recursively. In a separate analysis we show that by using "smoothed" probabilities instead, such short-lived indicators disappear. By using a smoothed estimated, however, would not be internally consistent with the forward looking nature of the index utility (1). In that matter, the merger wave probabilities reported in figure 6-8 fully reflect the agent's uncertainty about future path of deals flow.

5.2 Model Assessment

Figures 3-8 together support the idea that determinants of merger waves are heterogeneous in nature and specific for different industries. We now formally test which model specification has the highest in-sample explanatory power for the time series of deals flows. Figure 9 gives a first visual impression showing the model-implied intensity rates for each alternative set of regressors across industries. The solid red line represents the posterior median of $\lambda^m_{S_t,t}$ as specified in equation (7), which represents an estimate of the expected number of deals for z^m_{t-1} consistent with the behavioral hypothesis on merger waves. Except for Consumer Non-Durables, the performance of the model is remarkable. Except for Consumer Non-Durables abnormally increasing activity is captured in all industries.

[Insert Figure 9 about here]

The dashed black line shows the expected merger activity implied by the model estimated using the neoclassical hypothesis-related determinants. Posterior estimates are diverging from actual observations for Energy and Shops, underestimating merger activity in the second half of 2000s and between late 90s, beginning 2000s, respectively. The model fits reasonably well M&A activity for Money, Healthcare, Manufacturing and Other.[10]

Finally, the dash-dotted blue line shows the path of expected deals flow obtained by using aggregate macro-financial factors as unique regressors. Posterior estimates are essentially flat for Consumer Non-Durables, Energy and the Healthcare industry, whilst being excessively unstable for Shops when the actual number of deals increases. Overall, macro-financial factors show a remarkable performance in capturing the dynamics of merger activity as proxied by the number of deals. The visual impression provided by Figure 6-9, coupled with the statistical results in Figure 3-5, tend to show that the behavioral hypothesis proposed in Shleifer and Vishny 2003, Rhodes-Kropf and Viswanathan 2004, and Rhodes-Kropf et al. 2005 is likely more coherent with the deals flows.

We now formally test this result by comparing the (log) marginal likelihood for each model specification across industries. Marginal likelihoods allow to have a robust comparison across models as they take into account the latent nature of state indicators for

[10]It is worth to mention that there is no clear mapping between the posterior estimates of the switching parameters provided in Figure 3-5 and the results in Figure 9. As a matter of fact, while the posterior distributions of α^z_{12} answers to question "what determines a wave?", the model-implied intensity rates provide evidence of the goodness-of-fit produced by each alternative model specification. This also is a function of $\beta^m_{S_t}$, which estimates are available upon request from the authors.

merger waves and the uncertain nature of the model parameters. In that respect, they provide a measure of a model ability 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 computed as the harmonic mean of the conditional likelihood evaluated for each draw of the parameters from the full conditionals (see Gelfand and Dey 1994 and Newton and Raftery 1994). Table 4 shows the results;

[Insert Table 4 about here]

Except for Manufacturing, Shops and Money industries, the (log) marginal likelihood favors the behavioral hypothesis in explaining the dynamics of merger activity over the last thirty years. The higher goodness-of-fit provided by the neoclassical-related regressors for the Manufacturing industry is in-line with the posterior time series evidence provided in Figure 9. Overall, Table 4 confirms the visual impression above concerning the outperformance of a behavioral hypothesis on merger waves in explaining the dynamics of merger activity.

Figures 6-8 make clear that industry-specific merger waves are not coincident over time, regardless the preferred set of explanatory factors considered. One may argue this is a pure visual effect and does not necessarily imply any real disagreement across industry-specific paths. We now test formally the degree of synchrony across industries by computing a pairwise coincidence rate. More specifically, for each industry we construct a state indicator that takes value one for the industry j if the "filtered" probability of being in a merger wave is higher than a 0.5 threshold;

$$\mathbb{I}_{\{S_t^j=1\}} = \begin{cases} 1 \quad \text{if} \qquad p\left(S_t^j = 1 | \mathbf{y}_{1:t}^j, \mathbf{z}_{1:t}^j\right) \geq 0.5 \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases} \qquad \text{for} \qquad m = 1, \cdots, M, \tag{12}$$

with $\mathbf{y}_{1:t}^j, \mathbf{z}_{1:t}^j$ the time series of deals and determinants for the jth industry, respectively. Then we construct our coincidence measure between industry i and j, by simply counting in-sample the coincidence rate, namely how many times the indicator (12) gives the same signal between;

$$\mathbb{I}_{\left\{S_t^i = S_t^j\right\}} = \begin{cases} 1 & \text{if} \quad \mathbb{I}_{\left\{S_t^i = k\right\}} = \mathbb{I}_{\left\{S_t^j = k\right\}} \qquad \text{for} \qquad k = 1, 2, \\ 0 & \qquad \text{otherwise} \end{cases} \tag{13}$$

The coincidence rate between industry i and j can be finally constructed as;

$$\text{CoinRate}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\left\{S_t^i = S_t^j\right\}} \in [0, 1], \tag{14}$$

An in-sample rate equal to zero (one) implies that merger waves are perfectly disjoint (coincident). Figure 10 shows the heating maps summarizing the results for different model specifications. Darker values imply higher correlation among industry-specific merger waves.

[Insert Figure 10 about here]

Except few instances, cross-industry coincidence rates are below 0.6 for the neoclassical (left panel) and the behavioral (mid panel) hypothesis. Interestingly, the amount of convergence across industries increases if we use a common set of regressors. Indeed, by assuming the same macro-financial factors drive waves across industries the average rate of convergence is as high as 0.7, with peaks at 0.9 for the $\left(S_t^{Shops}, S_t^{NoDur}\right)$ pair. Although positive, such coincidence rates are sufficiently far from one to claim there is not perfect coordination of merger waves across industries. This is true regardless the model specification we use, with the macro-financial-related determinants showing a marginally higher cross-sectional convergence.

5.3 Out-of-Sample Prediction of Merger Activity

Further to understand the explanatory power of alternative theory-based determinants, in this section we investigate the predictive power of different theories in the context of our Markov regime-switching Poisson regression model with endogenous transition distributions, for each industry and for different forecasting horizons. Although key for corporate management and regulation, the predictability of merger activity did not receive much attention in the empirical corporate finance literature. Our modeling framework naturally embeds predictability. In general terms, for $h > 1$ the forecasting horizon, the predictive distribution from (7) given information available at time T, $\mathbf{z}_{1:T}, \mathbf{y}_{1:T}$ can be defined for the model $\mathcal{M}_i$ as

$$p\left(y_{T+h|T}|\mathbf{z}_{1:T}, \mathcal{M}_i\right) = \int \sum_{S_{T+h|T}} p\left(y_{T+h|T}|\mathbf{z}_{1:T}, \theta, S_{T+h|T}, \mathcal{M}_i\right) p\left(S_{T+h|T}, \theta|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \mathcal{M}_i\right) d\theta \tag{15}$$

with $y_{T+h|T}$ and $S_{T+h|T}$ representing the h-step ahead forecast for merger activity in the mth industry, and the latent state, respectively. Prediction of the latent state are generated by exploiting estimates of the transition probabilities at time T, $\xi_{lk,T}$ for $l, k = 1, 2$ as specified in the transitional dynamics (2). We approximate the predictive density (15) by using an importance sampling (IS) estimator (see, e.g. 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.[11] We simulate in each Gibbs step $y_{T+h|T}$ using (7) and the Markov regime-switching dynamics as data-generating process (DGP), where we replace the parameters and the in-sample latent variables by the draw from the posterior distribution.

We use the predictive density to obtain and evaluate one-month, one-quarter and one-year ahead forecasts for merger activities. We obtain the h-month ahead predictions through direct forecasting, as indirect procedures would entail iterating forward the

[11] 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 when computing the marginalized predictive likelihood by using the posterior distribution and for low dimensional prediction samples. The predictive density (15) can be alternatively approximated with an harmonic mean estimator; see Gelfand and Dey (1994). However, 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, e.g. Geweke 2005 for more details).

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 implies 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 $t_0 = 60$, and produce an h step ahead forecast $t_0 + h$ which is then evaluated against the observable deals flow at time $t_0 + h$. We then repeat this process until $T - h$, and will have time series of forecasts for $t = t_0, \cdots, T - h$, which we use to compute several prediction evaluation criteria.

As standard practice in the forecasting literature, we use the model-implied predictive density in order to compute a set of performance metrics. We first evaluate the point forecasts for different models using the square root of the mean squared forecast errors (RMSE). Gneiting (2011) showed that the RMSE is a consistent performance measure only when the point prediction equals the mean of the predictive distribution. As such, we base the RMSE of a model on the mean of the corresponding distribution of deals flow predictions.

Point forecasts, however, by definition cannot provide insight into the uncertainty that is associated with producing predictions. Merger activity density forecasts are more useful for this purpose, also by considering the whole predictive density we can carefully assess the ability of a model to predict unusual developments, such as sudden changes due to, for instance, unexpected valuation collapses, regulatory changes and alike. We compute the log predictive score based on the predictive density (15) for the model $\mathcal{M}_i$ as

$$LPS_h(\mathcal{M}_i) = \sum_{\tau=t_0}^{T-t_0-h} \log p\left(y_{\tau+h|\tau} | \mathbf{z}_{1:\tau}, \mathcal{M}_i\right), \tag{16}$$

Table 5 shows the results for different forecasting horizons, across different model specifications and for all the industries. Panel A shows the out-of-sample prediction for a short-term forecasting horizon, i.e. $h = 1$. Except for Consumer Non-Durables and Other, a model consistent with the behavioral hypothesis shows a lower RMSE with respect to the neoclassical specification, with an outperformance 18%, on average. Similarly, as far as macro-financial factors are concerned, industry-specific financial variables helps to predict better the one-step ahead merger activity in two-third of the times, with an average gain, in terms of squared error, equal to 21%. Evidences from the log-predictive score, however, make clear that, at least for short predictive horizons, macro-financial variables tend to underperform in predicting future deals flows. As a whole, the evidence for a short forecasting horizon is relatively mixed and tend to favor financial variables as useful predictors for merger activity.

[Insert Table 5 about here]

Panel B shows that, by increasing the predictive horizon, the power of behavioral-related determinants tends to be clearer. As we would expect the RMSE consistently increases across models and industries. Market-related variables now show the highest predictive power for six out of nine industries. The average gain with respect to competing model specifications is about 23% lower RMSE. Yet, the RMSE is comparable with competing regression specifications as the RMSE is only around 3% higher than the one obtained by using industry-specific economic shocks as explanatory variables. Log predictive scores provide again mixing evidences, although differences are somehow marginal, with the exceptions of Business Equipment, Shops and Money. While for short-to-mid horizons the gap between predictive performances is limited, Panel C shows that by increasing to $h = 12$ the difference across models becomes more sensible. The RMSE obtained from a neoclassical-related predictive regression is sensibly higher than using market-timing determinants, with the only exception of the Healthcare industry. Along the same intuition, industry-specific financial variables turn out to better capture not only the expected intensity of M&A activity but also the conditional distribution. Except for Business Equipment, Money and Other, the log-predictive score of behavioral-related models is higher than the competing specifications.

6 Conclusion

Methodologically, we propose a novel Markov regime-switching Poisson regression model with endogenous transition distributions to rationalize wave-like patterns in the intensity rate of industry-specific merger activity. Such an approach allows to acknowledge the endogeneity of commonly used regressors in determining the switching dynamics towards of abnormally high merger activity, i.e. a merger wave. Empirically, we show that merger waves varies significantly across industries, in terms of both timing and persistence. This suggests that any inference on existing economically justified competing explanations of merger waves would suffer form a generalization at the aggregate market level, as the observed cross-industry heterogeneity in waves is shown to be the consequence of different responses across industries to common or distinct drivers of merger activity.

References

Ahren, K., and J. Harford. 2014. The Importance of Industry Links in Merger Waves. *The Journal of Finance* 69:527–576.

Amisano, G., and G. Fagan. 2013. Money Growth and Inflation: A Regime Switching Approach. *Journal of International Money and Finance* 33:118–145.

Andrade, G., M. L. Mitchell, and E. Stafford. 2001. New Evidence and Perspectives on Mergers .

Barkoulas, J., C. Baum, and A. Chakraborty. 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.

Burda, M., M. Harding, and J. Hausman. 2011. Understanding Choice Intensity: A Poisson Mixture Model with Logit-Based Random Utility Selective Mixing. *Working Paper* .

Carter, C., and R. Kohn. 1994. On Gibbs sampling for state-space models. *Biometrika* pp. 541–553.

Casella, G., and C. P. Robert. 2004. *Monte Carlo Statistical Methods.* New York: Springer Verlag.

Chib, S. 1995. Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 432:1313–1321.

Choi, S., and B. Jeon. 2011. The Impact of the Macroeconomic Environment on Merger Activity: Evidence From US Time-Series Data. *Applied Financial Economics* 21:233–249.

Clark, T., and T. Davig. 2011. Decomposing the declining volatility of long-term inflation expectations. *Journal of Economic Dynamics and Control* 35:981–999.

Coase, R. 1937. The Nature of the Firm. *Economica* pp. 386–405.

Dube, J., G. Hitsch, and P. Rossi. 2009. Do Switching Costs Make Markets Less Competitive? *Journal of Marketing Research* 46:435–445.

Filardo, A. J. 1994. Business-Cycle Phases and Their Transitional Dynamics. *Journal of Business & Economic Statistics* 12:299–308.

Filardo, A. J., and S. F. Gordon. 1998. Business Cycle Durations. *Journal of Econometrics* 85:99–123.

Floegel, V., L. Johanning, and T. Gebken. 2005. The Dynamics Within Merger Waves: Evidence from Industry Merger Waves of the 1990s. *Working Paper* .

Frühwirth-Schnatter, S. 1994. Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis* 15:183–202.

Frühwirth-Schnatter, S., and R. Frühwirth. 2007. Auxiliary Mixture Sampling with Applications to Logistic Models. *Computational Statistics and Data Analysis* 51:3509–3528.

Frühwirth-Schnatter, S., and H. Wagner. 2006. Auxiliary Mixture Sampling for Parameter-Driven Models of Time Series of Counts with Applications to State Space Modeling. *Biometrika* 93:827–841.

Gärtner, D., and D. Halbheer. 2009. Are There Waves in Merger Activity After All? *International Journal of Industrial Organization* 27:708–718.

Gelfand, A. E., and D. K. Dey. 1994. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 501–514.

Geman, S., and D. Geman. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions* 6:721–741.

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*. Hoboken: John Wiley.

Geweke, J., and G. Amisano. 2012. Prediction and Mispecified Models. *American Economic Review* 102:482–486.

Gneiting, T. 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106:746–762.

Golbe, D., and L. White. 1988. A Time Series Analysis of Merger and Acquisitions in the U.S. Economy. *In: Auerbach, A. (Ed.), Corporate Takeovers: Causes and Consequences. NBER: University of Chicago Press, Chicago* pp. 265–302.

Golbe, D., and L. White. 1993. Catch a Wave: The Time Series Behavior of Mergers. *The Review of Economics and Statistics* pp. 493–499.

Gort, M. 1969. An Economic Disturbance Theory of Mergers. *The Quarterly Journal of Economics* pp. 624–642.

Groen, J., R. Paap, and F. Ravazzolo. 2013. Real-time Inflation Forecasting in a Changing World. *Journal of Business and Economic Statistics* 31:29–44.

Gugler, K., D. Mueller, and M. Weichselbaumer. 2012. The Determinants of Merger Waves: An International Perspective. *International Journal of Industrial Organization* 30:1–15.

Gugler, K., D. Mueller, and B. Yurtoglu. 2006. The Determinants of Merger Waves. *Working Paper* .

Guimarães, P., F. Octavio, and D. Woodward. 2003. A Tractable Approach to the Firm Location Decision Problem. *Review of Economics and Statistics* 85:201–204.

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.

Holmes, C. C., and L. Held. 2006. Bayesian Auxiliary Variable Models for Binary and Multinomial Regression. *Bayesian Analysis* 1:145–168.

Holmstrom, B., and S. N. Kaplan. 2001. Corporate Governance and Merger Activity in the US: Making Sense of the 1980s and 1990s. Tech. rep., National Bureau of Economic Research, Working Paper.

Jovanovic, B., and P. L. Rousseau. 2002. The Q-theory of Mergers. Tech. rep., National Bureau of Economic Research, Working Paper.

Justiniano, A., and G. Primiceri. 2008. The Time-Varying Volatility of Macroeconomic Fluctuations. *American Economic Review* 98:604–641.

Kass, R., and A. Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association* 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.

Kim, C.-J., J. Piger, and R. Startz. 2008. Estimation of Markov Regime-Switching Regression Models with Endogenous Switching. *Journal of Econometrics* 143:263–273.

Kim, S., N. Shepard, and S. Chib. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65:361–393.

Maksimovic, V., and G. Phillips. 2001. The Market for Corporate Assets: Who Engages in Mergers and Asset Sales and are There Efficiency Gains? *The Journal of Finance* 56:2019–2065.

Maksimovic, V., G. Phillips, and L. Yang. 2013. Private and Public Merger Waves. *The Journal of Finance* 68:2177–2217.

Martynova, M., and L. Renneboog. 2009. What Determines the Financing Decision in Corporate Takeovers: Cost of capital, Agency Problems, or the Means of Payment? *Journal of Corporate Finance* 15:290–315.

McFadden, D. 1974. *Conditional Logit Analysis of Qualitative Choice Behavior*. Frontiers of econometrics ed. New York: Academic Press.

Melicher, R., J. Ledolter, and L. D'Antonio. 1983. A Time Series Analysis of Aggregate Merger Activity. *Review of Economics and Statistics* 65:423–429.

Meligkotsidou, L., and P. Dellaportas. 2011. Forecasting with Non-Homogeneous Hidden Markov Models. *Statistics and Computing* 21:439–449.

Mitchell, M. L., and J. H. Mulherin. 1996. The Impact of Industry Shocks on Takeover and Restructuring Activity. *Journal of Financial Economics* 41:193–229.

Mulherin, J. H., and A. Boone. 2000. Comparing Acquisitions and Divestitures. *Journal of Corporate Finance* 6:117–139.

Newton, M., and A. Raftery. 1994. Approximate Bayesian Inference by the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society Series B. v56* 56:1–48.

Omori, Y., S. Chib, N. Shepard, and J. Nakajima. 2007. Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics* 140:425–449.

Primiceri, G. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* 72:821–852.

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

Resende, M. 1996. Mergers and Acquisitions in the UK: A Disaggregated Analysis. *Applied Economics Letters* 3:637–640.

Resende, M. 2008. Mergers and Acquisitions Waves in the UK: A Markov-Switching Approach. *Applied Financial Economics* 18:1067–1074.

Rhodes-Kropf, M., D. T. Robinson, and S. Viswanathan. 2005. Valuation Waves and Merger Activity: The Empirical Evidence. *Journal of Financial Economics* 77:561–603.

Rhodes-Kropf, M., and S. Viswanathan. 2004. Market Valuation and Merger Waves. *The Journal of Finance* 59:2685–2718.

Roberts, G., and S. Sahu. 1997. Updating schemes, covariance structure, blocking and parametrisation for the Gibbs sampler. *Journal of the Royal Statistical Society* 59:291 – 318.

Shleifer, A., and R. Vishny. 1992. Liquidation Values and Debt Capacity. *Journal of Finance* 32:337–347.

Shleifer, A., and R. W. Vishny. 2003. Stock Market Driven Acquisitions. *Journal of Financial Economics* 70:295–311.

Shugart, W., and R. Tollison. 1984. The Random Character of Mergers. *RAND Journal of Economics* 15:500–509.

Tanner, A., and W. Wong. 1987. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 82:528–540.

Terasvirta, T., and H. M. Anderson. 1992. Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics* 7:119–136.

Town, R. 1992. Merger Waves and the Structure of Merger and Acquisition Time-Series. *Journal of Applied Econometrics* 7:83–100.

Viscusi, W., J. W. Harrington, and J. Vernon. 2005. *Economics of Regulation and Antitrust*. Fourth ed. ed. Cambridge: MIT Press.

Appendix

A Gibbs Sampler

The practical implementation for non-linear models such as (7) requires the use of a Metropolis-Hastings steps to estimate the parameters vector. In this paper, we follow Frühwirth-Schnatter and Wagner (2006), and Frühwirth-Schnatter and Frühwirth (2007) 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 two sequences of auxiliary latent variables through data augmentation. The first of these sequences is the unobservable inter-arrival times of suitably chosen Poisson processes. This helps to get rid of the non-linearity that characterize the regression equation. However, these inter-arrival time introduce non-normality. We approximate such non-normality of the error term through a mixture of Gaussian distributions (see e.g., Kim, Shepard, and Chib 1998, Omori, Chib, Shepard, and Nakajima 2007 and Frühwirth-Schnatter and Wagner 2006). The resulting model may be thought as a partially Gaussian model, and a standard Gibbs sampler becomes feasible. As far as the time-varying transition probabilities are concerned, we follow Kaufmann (2015) and use a similar data augmentation approach to estimate a multi-logit specification to identify the regimes. The advantage of introducing additional layers of auxiliary latent variables is that draws from the posterior distributions of all parameters, including those driving the latent state of merger waves, are obtained from standard full conditional distributions.

In this section we provide details of the Gibbs sampler we propose for the estimation of the Markov regime-switching Poisson regression model with time-varying transition probabilities outlined in the main Section 3. Let us denote with $\mathbf{y}_{s:t} = (\mathbf{y}_s, \ldots, \mathbf{y}_t)$, $s \leq t$, a collection of vectors $\mathbf{y}_u$. The collections of parameters are defined as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$, respectively, where $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\alpha}_k)$, $k = 1, \ldots, K$, are the state-specific parameters. Given the model structure (7), the completed data likelihood is

$$p\left(\mathbf{y}_{1:T}, \mathbf{s}_{1:T} | \mathbf{z}_{1:T}, \boldsymbol{\theta}\right) = \prod_{k,l=1}^{K} \left(\prod_{t=1}^{T} y_t! \right)^{-1} \exp\left(-\sum_{t=1}^{T} \exp\left(\beta_t' \mathbf{z}_{t-1}\right) + \sum_{t=1}^{T} y_t \beta_t' \mathbf{z}_{t-1} \right) \xi_{lk,t}^{N_{lk,t}} \tag{A.1}$$

with $N_{lk,t} = \mathbb{I}_{\{l\}}\left(s_{t-1}\right) \mathbb{I}_{\{k\}}\left(s_t\right)$ and $\beta_t = \sum_{k=1}^{K} \beta_k \mathbb{I}_{\{k\}}(s_t)$. Combining the prior specifications (8)-(9) with the complete likelihood (A.1), we obtain the posterior density

$$p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T} | \mathbf{y}_{1:T}, \mathbf{z}_{1:T}\right) \propto p\left(\mathbf{y}_{1:T}, \mathbf{s}_{1:T} | \boldsymbol{\theta}, \mathbf{z}_{1:T}\right) p(\boldsymbol{\theta}) \tag{A.2}$$

Since the joint posterior distribution is not tractable the Bayesian estimator of the parameters cannot be obtained in analytical form. The random draws from the joint posterior distributions are obtained through a Gibbs sampler algorithm (Geman and Geman 1984). We propose a collapsed multi-move Gibbs sampling algorithm (see e.g. Roberts and Sahu 1997 and Casella and Robert 2004), where the hidden states and the parameter are sampled in blocks. More deeply, we combine a forward filtering backward sampling scheme (see Frühwirth-Schnatter 1994 and Carter and Kohn 1994 for more details) with the data augmentation technique proposed in Frühwirth-Schnatter and Wagner (2006), which makes feasible to sample the structural parameters. Although $\log \lambda_t$ in (7) is linear conditional on the states $\mathbf{S}_{1:T}$ and the parameters $\boldsymbol{\theta}$, the presence of the Poisson distribution causes non-normality as well as non-linearity of the mean λ_t in $\mathbf{S}_{1:T}$ and $\boldsymbol{\theta}$. Following Frühwirth-Schnatter and Wagner (2006) we design an approximate sampling scheme which helps to get rid of both non-normalities and non-linearities in an efficient way. At each iteration the Gibbs sampler sequentially cycles through the following steps:

Step 1. Draw the hidden interarrival times $\boldsymbol{\tau}$ conditional on $\boldsymbol{\theta}$ and $\mathbf{y}_{1:T}$.

Step 2. Draw the mixtures indicators $\mathbf{R}$ conditional on $\boldsymbol{\tau}$, $\mathbf{s}_{1:T}$, and $\boldsymbol{\beta}_k$.

Step 3. Draw $\mathbf{s}_{1:T}$ conditional on $\boldsymbol{\theta}, \boldsymbol{\tau}$ and $\mathbf{R}$.

Step 4. Draw $\boldsymbol{\beta}_k$ conditional on $\boldsymbol{\tau}$, $\mathbf{s}_{1:T}$ and $\mathbf{R}$.

Step 5. Draw the latent state utilities $\mathbf{S}^{\star}_{1:T}$ conditional on $\mathbf{s}_{1:T}$ and $\boldsymbol{\alpha}_k$.

Step 6. Draw the mixtures indicators $\mathbf{R}^s$ conditional on $\mathbf{S}^{\star}_{1:T}$, and $\boldsymbol{\alpha}_k$.

Step 7. Draw the parameters $\boldsymbol{\alpha}_k$ conditional on $\mathbf{S}^{\star}_{1:T}$, and $\mathbf{R}^s$.

Step 1 and 2 are implemented to get rid of the non-linearity and non-normality that characterize the Poisson regression. Step 5 and 6 are used to estimate the multi-logit by which the time-varying transition probabilities are characterized. In the following we provide a detailed description of each step of the Gibbs sampler.

Step 1. Sampling the Hidden Interarrival Times τ

For each 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 occured in the time period $[0,\ 1]$. The first step of the Gibbs sampler is to create such a Poisson process for each y_t, for $t = 1, ..., T$, and introduce such interarrival times τ_{tj}, for $j = 1, ..., (y_t + 1)$, as missing data. We start from the fact in a Poisson process the interarrival times are distributed as Exponential random variables, $\tau_{tj} \sim Exp\left(\lambda_t\right)$, such that

$$\tau_{tj}|\boldsymbol{\theta}, s_t = \frac{\xi_{tj}}{\lambda_t}, \qquad \xi_{tj} \sim Exp\left(1\right), \tag{A.3}$$

This can be rewritten as a linear model of the form

$$-\log \tau_{tj}|\boldsymbol{\theta}, s_t = \beta_t' \mathbf{z}_{t-1} + \epsilon_{tj}, \qquad \epsilon_{tj} = -\log \xi_{tj}, \qquad \xi_{tj} \sim Exp\left(1\right), \tag{A.4}$$

Let $\boldsymbol{\tau} = \left(\tau_{tj}, j = 1, ..., (y_t + 1), t = 1, ..., T\right)$ denote the collection of all interarrival times, the full conditional distribution $p\left(\mathbf{s}_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}_{1:T}\right)$ depends on $\mathbf{y}_{1:T}$ only through $\boldsymbol{\tau}$; $p\left(\mathbf{s}_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}_{1:T}\right) = p\left(\mathbf{s}_{1:T}|\boldsymbol{\theta}, \boldsymbol{\tau}\right)$. Interarrival times are independent for different time points t, given $\boldsymbol{\theta}, \mathbf{s}_{1:T}$ and $\mathbf{y}_{1:T}$;

$$p\left(\boldsymbol{\tau}|\boldsymbol{\theta}, \mathbf{s}_{1:T}, \mathbf{y}_{1:T}\right) = \prod_{t=1}^{T} p\left(\tau_{t1}, ..., \tau_{t,n+1}|y_t, \boldsymbol{\theta}, \mathbf{s}_{1:T}\right), \tag{A.5}$$

where $N = y_t + 1$. For fixed t, the interarrival times are stochastically dependent, and the joint distribution factorises as

$$p\left(\tau_{t1}, ..., \tau_{t,N}, \tau_{t,N+1}|y_t = N, \boldsymbol{\theta}, \mathbf{s}_{1:T}\right) = p\left(\tau_{t,n+1}|y_t = N, \boldsymbol{\theta}, \mathbf{s}_{1:T}, \tau_{t1}, ..., \tau_{t,N}\right) p\left(\tau_{t1}, ..., \tau_{t,N}|y_t = N\right), \tag{A.6}$$

The first N interarrival times are independent of all model parameters, and are determined only by the observed number of counts y_t. By well-known 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 $p(\tau_{t1},...,\tau_{t,n}|y_t = N)$ is approximated sampling the order statistics $u_{t(1)},...,u_{t(N)}$ of $N = y_t$ $Uniform[0,1]$ random variables, and define the interarrival times as their increments $\tau_{tj} = u_{t(j)} - u_{t(j-1)}$, for $j = 1,...,N$, where $u_{t(0)} = 0$. Only the final interarrival 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. Conditionally 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 on each other, the waiting time after 1 is then distributed as $Exp(\lambda_t)$. Therefore, the final arrival time is sampled from $p(\tau_{t,N+1}|y_t = N, \boldsymbol{\theta}, \mathbf{s}_{1:T}, \tau_{t1},...,\tau_{t,n})$ by defining $\tau_{t,N+1} = 1 - \sum_{j=1}^{N} \tau_{tj} + \xi_t$, where $\xi_t \sim Exp(\lambda_t)$.

Step 2. Sampling the Indicators R for the Mixture Approximation

The linear approximation of Step 1 implies an error term

$$\epsilon_{tj} = -\log \xi_{tj}, \qquad \xi_{tj} \sim Exp(1) \tag{A.7}$$

the density of which is defined as

$$p(\epsilon) = \exp(-\epsilon - \exp(-\epsilon)),$$

Now to obtain a model that is conditionally Gaussian we approximate this distribution by a mixture of R normal components:

$$p(\epsilon) = \exp(-\epsilon - \exp(-\epsilon)) \approx \sum_{r=1}^{R} \omega_r N\left(\epsilon; m_r, \sigma_r^2\right), \tag{A.8}$$

where, for $r = 1,...,R$, m_r and σ_r^2 are the mean and variance of the $N(\epsilon;.,.)$ Gaussian density. We follow Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter and Frühwirth (2007), and use $R = 10$ components. Let $\mathbf{R} = (r_{tj}, j = 1,...,(y_t+1), t = 1,...,T)$ denote the collection of latent component indicators. Conditional on $\boldsymbol{\tau}$ and $\mathbf{R}$, the non-normal, non-linear model (7) reduces to a linear, Gaussian model such as

$$-\log \tau_{tj}|\boldsymbol{\theta}, s_t, r_{tj} = \beta_t' \mathbf{z}_{t-1} + m_{r_{tj}} + \epsilon_{tj}, \qquad \epsilon_{tj}|r_{tj} \sim N\left(0, \sigma_{r_{tj}}^2\right), \tag{A.9}$$

To sample $\mathbf{R}$ we can use the fact that all indicators are conditional independent given $\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{s}_{1:T}$ and $\mathbf{y}_{1:T}$;

$$p(\mathbf{R}|\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{s}_{1:T}, \mathbf{y}_{1:T}) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} p(r_{tj}|\tau_{tj}, \boldsymbol{\theta}, s_t), \tag{A.10}$$

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

$$p(r_{tj} = r|\tau_{tj}, \boldsymbol{\theta}, s_t) \propto p(\tau_{tj}|r_{tj} = r, \boldsymbol{\theta}, s_t)\, \omega_r, \tag{A.11}$$

where

$$p\left(\tau_{tj}|r_{tj}=r,\boldsymbol{\theta},s_t\right)\propto\frac{1}{\sigma_r}\exp\left(-\frac{1}{2\sigma_r^2}\left(-\log\tau_{tj}-\beta_t'\mathbf{z}_{t-1}-m_r\right)^2\right),\tag{A.12}$$

Step 3. Sampling $\mathbf{s}_{1:T}$

Conditional on the hidden interarrival times $\boldsymbol{\tau}$, the mixture component indicators $\mathbf{R}$, and the parameters $\boldsymbol{\theta}=(\beta,\alpha)$, the observation equation takes the form

$$-\log\tau_{tj}|\boldsymbol{\theta},s_t,r_{tj}=\beta_t'\mathbf{z}_{t-1}+m_{r_{tj}}+\epsilon_{tj},\qquad\epsilon_{tj}|r_{tj}\sim N\left(0,\sigma^2_{r_{tj}}\right),\tag{A.13}$$

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

$$\tilde{\mathbf{y}}_t=\begin{pmatrix}-\log\tau_{t1}-m_{r_{t1}}\\ \vdots\\ -\log\tau_{tN}-m_{r_{tN}}\end{pmatrix},\tag{A.14}$$

Then we can rewrite an augmented model in a state-space form as

$$\tilde{\mathbf{y}}_t=\beta_t'\tilde{\mathbf{z}}_{t-1}+\eta_t,\qquad\eta_t\sim N\left(0,\Sigma_t\right),\tag{A.15}$$

where $\Sigma_t=diag\left(\sigma^2_{r_{t1}},...,\sigma^2_{r_{tN}}\right)$, and $\tilde{\mathbf{z}}_{t-1}$ contains N rows of $\mathbf{z}_{t-1}$, i.e.

$$\tilde{\mathbf{z}}_{t-1}=\begin{pmatrix}\mathbf{z}_{t-1}'\\ \vdots\\ \mathbf{z}_{t-1}'\end{pmatrix},$$

Now we have a state-space model for repeated measurement in which the transition equation for the hidden states $\mathbf{s}_{1:T}$ is the same as for the original Poisson regression model (7). In fact, (A.15) can be treated as a standard Markov regime-switching seemingly unrelated regression model. Thus, we can use a forward filtering backward sampling algorithm (see Frühwirth-Schnatter 1994 and Carter and Kohn 1994). As the state s_t is discrete valued the FFBS is applied in its Hamilton form. The Hamilton filter iterates in two steps: The prediction step at each time t is

$$p\left(s_{t+1}=j|\boldsymbol{\theta},\tilde{\mathbf{y}}_{1:t}\right)=\sum_{k=1}^{K}\xi_{kj,t}p\left(s_t=k|\boldsymbol{\theta},\tilde{\mathbf{y}}_{1:t}\right)\tag{A.16}$$

The updating step can be easily derived as

$$p(s_{t+1}=k|\boldsymbol{\theta},\tilde{\mathbf{y}}_{1:t+1})=\frac{p(\tilde{\mathbf{y}}_{t+1}|s_{t+1}=k,\boldsymbol{\theta},\tilde{\mathbf{z}}_t)p(s_{t+1}=k|\tilde{\mathbf{y}}_{1:t},\boldsymbol{\theta})}{p\left(\tilde{\mathbf{y}}_{t+1}|,\tilde{\mathbf{z}}_t,\boldsymbol{\theta}\right)}\tag{A.17}$$

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

$$p\left(\tilde{\mathbf{y}}_{t+1}|\tilde{\mathbf{z}}_t,\boldsymbol{\theta}\right)=\sum_{k=1}^{K}p\left(\tilde{\mathbf{y}}_{t+1}|s_{t+1}=k,\boldsymbol{\theta},\tilde{\mathbf{z}}_t\right)p\left(s_{t+1}=k|\boldsymbol{\theta},\tilde{\mathbf{y}}_{1:t}\right)\tag{A.18}$$

The draw $p(\mathbf{s}_{1:T}|\mathbf{y}_{1:T},\boldsymbol{\theta})$ can then be obtained recursively and backward in time by using the smoothed probabilities as

$$p(\mathbf{s}_{1:T}|\mathbf{y}_{1:T},\boldsymbol{\theta}) = p\left(s_T|\mathbf{y}_{1:T},\boldsymbol{\theta}\right) \times \prod_{t=1}^{T-1} p\left(s_t|s_{t+1},\mathbf{y}_{1:t},\boldsymbol{\theta}\right) \tag{A.19}$$

where for instance

$$p\left(s_t = k|s_{t+1} = j, \mathbf{y}_{1:t},\boldsymbol{\theta}\right) = \frac{\xi_{kj,t}\, p\left(s_t = k|\mathbf{y}_{1:t},\boldsymbol{\theta}\right)}{p\left(s_{t+1} = j|\theta, \mathbf{y}_{1:t}\right)} \tag{A.20}$$

Step 4. Sampling $\boldsymbol{\beta}_k$

Conditional on $\mathbf{s}_{1:T}, \mathbf{y}_{1:T}$ and the Gaussian linear approximation (A.15), the prior (8) are conjugate. Let $\mathcal{T}_k = \{t : s_t = k\}$, the posterior for the regime-dependent betas β_k is

$$(\boldsymbol{\beta}_k|\mathbf{y}_{1:T},\mathbf{s}_{1:T}) \sim \mathcal{N}_p\left(B_k^*\left(\sum_{t\in\mathcal{T}_k} \tilde{\mathbf{z}}'_{t-1}\tilde{\mathbf{y}}_t + B_k^{-1}b_k\right), B_k^*\right) \tag{A.21}$$

with $B_k^* = \left(\sum_{t\in\mathcal{T}_k} \tilde{\mathbf{z}}_{t-1}\tilde{\mathbf{z}}'_{t-1} + B_k^{-1}\right)^{-1}$ the posterior scale parameter. Notice the conditional variance structure Σ_t is given by the latent indicators $\mathbf{R}$. As such, given $\mathbf{R}$ the betas are independent on Σ_t.

Step 5. Sampling the Latent State Utilities $\mathbf{S}^{\star}_{1:T}$

The sampling of the parameters governing the transition probabilities of the Markov regime switching process are drawn based on a data augmentation process as in Kaufmann (2015). I a first step, we extend the model to a non-normal specification for so-called state-dependent latent utilities;

$$\begin{aligned} S^u_{k,t} &= Z'_t\alpha_k + \nu_{k,t}, \qquad \forall k \in \mathcal{K}_{-k_0}, \\ S^u_{k_0,t} &= \nu_{k_0,t}, \qquad \text{for the identification restriction of the reference state} \qquad \alpha_{k_0} = 0, \end{aligned} \tag{A.22}$$

where $Z_t = \left(z_t D^{(1)}_{t-1}, z_t D^{(2)}_{t-1}, \cdots, z_t D^{(K)}_{t-1}, D^{(1)}_{t-1}, D^{(2)}_{t-1}, \cdots, D^{(K)}_{t-1}\right)$ with $D^{(k)}_{t-1}$ a dummy variable that takes value equal to one if $S_{t-1} = k$. As for Step 1, the errors $\nu_{k,t}$ are i.i.d. over t, and follow a type I extreme value distribution. Conditional on the latent state variables $\mathbf{s}_{1:T}$, to sample $S^u_{k,t}$, for each $t = 1, \cdots, T$, we sample K independent random numbers $V_{1t}, \cdots, V_{Kt}$ from a uniform $Uniform[0, 1]$ and obtain:

$$S^u_{k,t} = -\log\left(-\frac{\log\left(V_{1t}\right)}{\sum_{k=1}^K \tilde{\lambda}_{kt}} - \frac{\log\left(V_{kt}\right)}{\tilde{\lambda}_{kt}}\mathbb{I}_{\{s_t\neq k\}}\right), \tag{A.23}$$

with $\tilde{\lambda}_{kt} = \exp\left(Z'_t\alpha_k\right)$. The simulation step (A.23) is derived by exploiting the assumption that the maximal value of $S^u_{j,t}$ is obtained in correspondence of the observed state. As such, $\exp\left(-S^u_{j,t}\right)$ is the minimum value among all of the possible alternatives $\exp\left(-S^u_{k,t}\right)$ if $s_t = j$. The type I extreme value distribution of $\nu_{k,t}$ implies that $\exp\left(-S^u_{k,t}\right) \sim Exp\left(\tilde{\lambda}_{kt}\right)$. By using the same idea of sampling the hidden inter-arrival times, given the minimum, all other state-dependent auxiliary utilities are conditionally independent, as implied by (A.23).

Step 6. Sampling the Indicators $\mathbf{R}^s$ for the Mixture Approximation

Conditional on the latent state-utilities S^u_{kt}, we can sample a component indicator which is used to obtain a Gaussian approximation of the initial non-Gaussian (A.22). Let $\mathbf{R}^s = \left(r^s_{t,k}, k = 1, \cdots, K, t = 1, \cdots, T\right)$ denote the collection of latent component indicators. Conditional on $S^u_{1:T}$ and $\mathbf{R}^s$ the non-normal, non-linear model (A.22) reduces to

$$S^u_{k,t} = Z_t'\alpha_k + \tilde{m}_{r^s_{t,k}} + \epsilon_{t,k}, \qquad \epsilon_{t,k}|r^s_{t,k} \sim N\left(0, \tilde{\sigma}^2_{r^s_{t,k}}\right),$$

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

$$p\left(r^s_{t,k} = r|S^u_{k,t}, \alpha_k\right) \propto p\left(S^u_{k,t}|r^s_{t,k} = r, \alpha_k, s_{1:T}\right)\omega^s_r,$$

where

$$p\left(S^u_{k,t}|r^s_{t,k} = r, \alpha_k, s_{1:T}\right) \propto \frac{1}{\tilde{\sigma}^2_r}\exp\left(-\frac{1}{2\tilde{\sigma}^2_r}\left(S^u_{k,t} - Z_t'\alpha_k - \tilde{m}_r\right)^2\right), \qquad k \in \mathcal{K}_{-k_0},$$

where $r = 1, \cdots, 10$ and the respective $\tilde{m}_r$, $\tilde{\sigma}^2_r$ and $\tilde{\omega}_r$ are taken from Frühwirth-Schnatter and Frühwirth (2007).

Step 7. Sampling the Parameters $\boldsymbol{\alpha}_k$

For all given state-utilities $S^u_{1:T} = \left(S^u_{1,1}, \cdots, S^u_{K,1}, \cdots, S^u_{K,T}\right)$ and all component indicators $\mathbf{R}^s = \left(r^s_{11}, \cdots, r^s_{K1}, \cdots, r^s_{KT}\right)$, 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 Gaussian prior structure (9), the posterior distribution is updated as

$$(\alpha_k|\mathbf{y}_{1:T}, S^u_{1:T}, \mathbf{R}^s) \sim \mathcal{N}_N\left(A^*_k\left(\sum_{t=1}^T Z_t\left(S^u_{k,t} - \tilde{m}_{r^s_{t,k}}\right)/\tilde{\sigma}^2_{r^s_{t,k}} + A^{-1}_k a_k\right), A^*_k\right), \tag{A.24}$$

with $A^*_k = \left(\sum_{t=1}^T Z_t Z_t'/\tilde{\sigma}^2_{r^s_{t,k}} + A^{-1}_k\right)^{-1}$ the posterior scale parameter.

B Testing the Number of Regimes

Our random utility discrete choice model specification outlined in Section 3 can be generalized to have more than two regimes. In this section we test the null hypothesis of $\mathcal{H}_0 : K = 2$ against the alternative $\mathcal{H}_1 : K > 2$. As usual in the Bayesian literature we based hypothesis testing on Bayes factors comparing the model with two regimes $\mathcal{M}_2$ against a model with three regimes $\mathcal{M}_3$. Bayes factors are based on marginal likelihoods (see Kass and Raftery 1995); comparing, the two-state vs. the three-state model can be accomplished by computing

$$\mathcal{B}_{2,3} = \frac{p(\mathbf{y}_{1:T}|\mathbf{z}_{1:T}; \mathcal{M}_2)p\left(\mathcal{M}_2\right)}{p(\mathbf{y}_{1:T}|\mathbf{z}_{1:T}; \mathcal{M}_3)p\left(\mathcal{M}_3\right)}, \tag{B.25}$$

for each of the M industries. Marginal likelihoods are computed by integrating out both parameter and state uncertainty;

$$p(\mathbf{y}_{1:T}|\mathbf{z}_{1:T};\mathcal{M}_i) = \int \sum_{S} p\left(\mathbf{y}_{1:T}|\boldsymbol{\theta}, \mathbf{s}_{1:T}, \mathbf{z}_{1:T};\mathcal{M}_i\right) p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}|\mathbf{y}_{1:T}, \mathbf{z}_{1:T-1};\mathcal{M}_i\right) d\boldsymbol{\theta}, \tag{B.26}$$

with S the set of states, and the posterior distribution $p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}|\mathbf{y}_{1:T}, \mathbf{z}_{1:T}\right)$ obtained from the Gibbs sampler. The marginal likelihood however, is not available in closed form and must be approximated numerically as in Chib (1995). Table B.1 shows the marginal likelihoods and corresponding Bayes factors (with $p\left(\mathcal{M}_2\right) = p\left(\mathcal{M}_3\right)$) across industries and using different sets of theory-based determinants.

[Insert Table B.1 about here]

Panel A shows the results obtained using the behavioral-related determinants. The marginal likelihood is decisively in favor of a model with two regimes. The third raw reports the corresponding Bayes factors in log scale. The evidence strongly points towards a model with two regimes vs. three states. Panel B and C show the results by using the set of neoclassical-related and macroeconomic determinants, respectively. Bayes factors confirm that, except for Utilities which offers borderline evidence, the empirical results are again in favor of $K = 2$. As a result, in the following we use the two-state model as our preferred specification.

C MCMC Convergence Analysis

We report the results of a convergence analysis of the MCMC sampler for the Poisson Markov Regime-Switching regression model outlined in Section 2 in the main text. We investigate convergence properties of our MCMC approach by computing a set of inefficiency factors and t-tests for equality of the means across sub-samples of the MCMC chain (see e.g. Geweke 1992, Primiceri 2005, Justiniano and Primiceri 2008, Clark and Davig 2011 and Groen, Paap, and Ravazzolo 2013). In order to assess inefficiencies we used as an example the model specification with the behavioral-related explanatory factors and $K = 2$, with $K = 1$ the state of no merger waves.

For each individual parameter and latent variable we measure an inefficient as $(1 + 2\sum_{f=1}^{\infty}\rho_f)$, where ρ_f is the f_{th} order auto-correlation of the chain of draws. This inefficiency factor tells how much information do we actually have about parameters and equals the variance of the mean of the posterior draws from the MCMC sampler, divided by the variance of the mean assuming independent draws. As such, if we require that the variance of the mean of the MCMC posterior draws should be limited to be at most 1% of the variation due to the data (measured by the posterior variance), the inefficiency factor provides an indication of the minimum number of MCMC draws to achieve this (see Kim et al. 1998). If there are some correlation between successive samples, then we might expect that our sample has not revealed as much information of the posterior distribution of our parameter as we could have gotten if the samples draws were independent. We compute the inefficiency factor for all model parameters and applied on a range of choices for the total number of posterior draws as well as burn-in period lengths and thinning for the Poisson Markov Regime-Switching regression specification. We consider the base case with two regressors plus an intercept and two underlying hidden states.

Tables C.1 provide a summary of the results showing that, for most parameters and latent variables, our MCMC sampler is very efficient and that it requires far less than 10000 retained posterior draws to be able to do a reasonably accurate inferential analysis.

[Insert Table C.1 about here]

We also compute the p-value of the Geweke (1992) t-test for the null hypothesis of equality of the means computed with the first 20 percent and last 20 percent of the sample of retained draws. For this particular convergence diagnostic test we compute the variances of the respective means using the Newey and West (1987) heteroskedasticity and autocorrelation robust variance estimator with a bandwidth set to 4% of the utilized sample sizes. Such convergence statistics is still computed for the complete Poisson Markov Regime-Switching regression specification estimated over the sample period 1983:01 - 2013:12. Table C.2 shows the results.

[Insert Table C.2 about here]

The convergence diagnostic tests in Table C.2 confirm the efficiency of the MCMC sampler we propose. For example, in the case of the **B** parameters the null hypothesis of equal means across sub-samples of the retained draws is hardly ever rejected at the 5% confidence interval. Thus, inference in our factor model appears to be reasonably accurate when we base posterior inference on 50000 draws with a burn-in of 10000 and thin value of 5. Such a choice of the number of draws keeps the computational burden relatively low, at the benefit of inference precision as shown in Table C.1 and Table C.2.

Table 1. Variables Description

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 computed as industry specific median absolute annual change, computed across all firms with data available on COMPUSTAT, on the basis their SIC codes and according to a 12-industry classification. Panel C reports the aggregate macroeconomic variables.

Panel A: Stock Valuation Variables (related to the behavioral hypothesis)			
	Description	Frequency	Source
M/B	Industry-specific market-to-book ratio	annual	Kenneth French's data
sd(M/B)	Standard deviation of the industry-specific market-to-book ratios	annual	COMPUSTAT
Returns	Industry-specific stock returns (value-weighted)	Monthly	Kenneth French's data
svar	Realized variance of aggregate market returns	Monthly	COMPUSTAT
dy	Aggregate dividend-yield	Monthly	COMPUSTAT

Panel B: Industry-specific economic shocks (related to the neoclassical hypothesis)			
	Description	Frequency	Source
Margin	Margin on sales: Net income/Sales.	annual	COMPUSTAT
AT	Asset turnover: Sales/Assets.	annual	COMPUSTAT
R&D	Research and development expenses scaled by assets at the beginning of the period	annual	COMPUSTAT
CAPX	Capital expenditures scaled by assets at the beginning of the period	annual	COMPUSTAT
Emp	Number of employees	annual	COMPUSTAT
ROA	Returns on assets: Net income/Assets at the beginning of the period	annual	COMPUSTAT
Sale	Sales	annual	COMPUSTAT
CL	Capital liquidity: spread between the average interest rate on commercial and industrial loans and the Fed Funds rate	quarterly	FRED
Reg	Dummy variable to indicate whether or not the industry has been subject to a de-regulatory event	Annual	Viscusi et al. (2005)
Margin	Margin on sales: Net income/Sales. We consider the industry specific median absolute change,	and Factiva	

Panel C: Macroeconomic variables			
	Description	Frequency	Source
Rec	NBER indicator for recessions from the Peak through the Trough	monthly	FRED
IP	Industrial production (year-on-year growth)	monthly	FRED
Term	Term spread: the yield spread between the 10-year government bond and the 3-month T-Bill	monthly	FRED
Credit	Credit spread: the yield spread between 20-year Baa and Aaa corporate bonds	monthly	FRED
Rf	Real risk free rate: difference between the 1-month T-Bill rate and the CPI inflation rate	monthly	FRED
Agg M/B	Aggregate market-to-book ratio	annual	Compustat

Table 2. Descriptive Statistics

The table reports a set of descriptive statistics for the number of deals across industries and the corresponding industry-specific regressors used in testing both the behavioural and the neoclassical hypothesis. Panel A reports sufficient statistics on the number of deals. The sample contains all announced bids by US private and public acquirers that were announced in the period from 1983 to 2014, for which the bidder did not previously own a majority interest in the target and is indeed seeking to obtain a majority interest through the transaction. Data on M&A deals flow are collected from Thomson One Banker. An observation is included in the final sample if: the transaction value is above \$5 million; the transaction is not a buyback, an exchange offer, a recapitalization, or an acquisition of partial or remaining interest. The final sample includes overall 60,305 observations for which the deal value is disclosed. Merger activity is aggregated at the industry level monthly, assigning each deal 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. Panel B reports the sample average values for the industry-specific regressors used in each Markov regime-switching Poisson regression specification. Industry-specific regressors are computed annually as median absolute changes of the corresponding variable and then transformed in monthly terms through a cubic spline approximation. The variables considered are cash-flow margin on sales (cash flow scaled by sales, *Marg*), asset turnover (sales divided by beginning-of-period assets, AT), return on assets (ROA), capital expenditures (scaled by beginning-of-period assets, $CAPX$), research and development (scaled by beginning-of-period assets, *R&D*), employee growth (*Emp*), sales growth (*Sale*). Panel C reports the industry-specific market-to-book ratio M/B and its cross-sectional standard deviation $sd(M/B)$, the industry-specific market returns (*Return*). The sample period is 1983:01-2014:12.

Panel A: Number of Deals (Aggregated Monthly)												
	NoDur	Durbl	Manuf	Enrgy	Chems	BusEq	Telcm	Utils	Shops	Hlth	Money	Other
Mean	6.065	2.154	10.617	7.964	2.323	21.495	8.865	3.745	8.932	10.195	56.781	17.901
Median	6.000	2.000	10.000	7.000	2.000	17.000	7.000	3.000	8.000	10.000	55.000	15.000
Std	3.198	1.645	5.242	4.506	1.698	16.396	7.119	2.608	5.395	5.473	23.832	10.095
Min	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	16.000	1.000
Max	19.000	8.000	36.000	34.000	9.000	98.000	35.000	17.000	31.000	31.000	152.000	64.000
AR(1)	0.338	0.349	0.550	0.378	0.192	0.906	0.734	0.349	0.715	0.553	0.828	0.753

Panel B: Industry Specific Economic Shocks (related to the Neoclassical hypothesis)												
	NoDur	Durbl	Manuf	Enrgy	Chems	BusEq	Telcm	Utils	Shops	Hlth	Money	Other
Margin	0.021	0.024	0.033	0.169	0.032	0.077	0.060	0.015	0.004	0.171	0.034	0.062
AT	0.024	0.027	0.030	0.148	0.033	0.082	0.063	0.013	0.014	0.160	0.039	0.052
ROA	0.035	0.043	0.041	0.065	0.041	0.090	0.037	0.006	0.030	0.109	0.008	0.057
CAPX	0.014	0.015	0.016	0.062	0.014	0.019	0.020	0.012	0.017	0.018	0.002	0.024
R&D	0.002	0.005	0.003	0.001	0.003	0.020	0.002	0.000	0.000	0.032	0.000	0.001
Emp	0.080	0.094	0.084	0.121	0.067	0.142	0.094	0.031	0.096	0.142	0.083	0.119
Sale	0.111	0.140	0.133	0.270	0.119	0.210	0.140	0.077	0.123	0.246	0.142	0.178

Panel C: Stock Valuation Variables (related to the Behavioral hypothesis)												
	NoDur	Durbl	Manuf	Enrgy	Chems	BusEq	Telcm	Utils	Shops	Hlth	Money	Other
M/B	0.372	0.848	0.540	0.692	0.446	0.369	0.695	1.017	0.412	0.283	0.738	0.556
sd(M/B)	0.161	0.136	0.149	0.140	0.143	0.139	0.142	0.132	0.149	0.157	0.146	0.108
Return	0.872	0.760	0.789	0.826	0.671	0.629	0.719	0.506	0.832	0.570	0.753	0.893

Table 3. Deregulatory Events

This table reports major de-regulatory initiatives during the sample period 1983:01-2014:12. 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	Enrgy
1991	Federal Deposit Insurance Corporation Improvement Act	Money
1992	Cable Television Consumer Protection and Competition Act	Other
	Energy Policy Act	Enrgy
	FERC Order 636	Utils
1993	Elimination of State regulation of cellular telephone rates	Telcm
	Negotiated Rates Act	Other
1994	Trucking Industry and Regulatory Reform Act	Other
	Interstate Banking and Branching Efficiency Act	Money
1995	Interstate Commerce commision termination act	Telcm
1996	Telecommuncations Act	Telcm
	FERC Order 888	Utils
1999	FERC Order 2000	Utils
	Gramm-Leach-Bliley Act	Money

Table 4. Marginal Likelihoods

Marginal likelihoods evidence. This table reports summary statistics about the (log) of the marginal likelihood computed from alternative specifications of our Markov regime switching Poisson regression model with time-varying transition probabilities. The marginal likelihood for each model is computed as the harmonic mean of the conditional likelihood evaluated for each draw of the parameters from the full conditionals (see Gelfand and Dey 1994 and Newton and Raftery 1994). Posterior estimates are based on 50,000 draws with a burn-in of 10,000 and thin value of 5. The goodness of fit of the model is assessed by using different theory-based regressors and for each industry. The sample period is 1983:01-2014:12, monthly. Boldfaced numbers indicates the highest fitting performance as indicated by a higher log marginal likelihood.

Panel A: Marginal likelihoods (log)

	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
Behavioral	**-2219.83**	-3397.67	**-3058.13**	**-7303.63**	**-3126.65**	-2701.69	**-3231.81**	-18208.61	**-5962.29**
Neoclassical	-2419.78	**-3358.81**	-3066.74	-7652.38	-3195.01	-2660.80	-4185.99	-19269.53	-6928.04
Macro-Financial	-2522.54	-3384.86	-3098.52	-7387.16	-3139.22	**-2259.81**	-3295.76	**-18106.92**	-5993.38

Table 5. Out-of-Sample Predictive Performances

Predicting merger waves over different horizons. This table reports two different performance measures to assess the predictive power of our Markov regime switching Poisson regression model with time-varying transition probabilities for different forecasting horizons and alternative specifications of regressors across industries. Panel A reports the results for $h = 1$, by showing the standard Root Mean Squared Errors (RMSE) and the Log Predictive Score, both computed sampling from the predictive density of the model obtained integrating out uncertainty on both parameters and latent states. Panel B and C shows the same performance metrics for $h = 3$ months and $h = 12$ months, respectively. The sample period is 1983:01-2014:12, monthly. Boldfaced numbers indicates the highest forecasting performance for a given metrics.

Panel A: Forecasting Horizon, $h = 1$									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
	RMSE								
Behavioral	0.686	**2.421**	**0.923**	1.574	**0.346**	**1.120**	0.955	**1.475**	1.819
Neoclassical	**0.619**	3.694	1.059	1.829	0.572	1.144	1.177	1.587	**1.661**
Macro-Financial	1.327	3.541	1.411	**1.453**	0.560	1.160	**0.786**	1.707	1.700
	Log-Predictive Score								
Behavioral	**-2.094**	-1.114	-2.177	**-3.466**	-2.539	-1.566	**-2.389**	-0.684	**-1.135**
Neoclassical	-2.356	**-1.102**	**-2.011**	-8.661	**-2.192**	**-1.373**	-2.524	**-0.277**	-1.77
Macro-Financial	-3.905	-2.143	-2.747	-3.592	-4.022	-3.876	-2.954	-1.450	-1.888
Panel B: Forecasting Horizon, $h = 3$									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
	RMSE								
Behavioral	**1.376**	**3.973**	**1.538**	**0.933**	2.530	2.588	3.909	**2.016**	**2.077**
Neoclassical	1.606	5.169	1.675	2.891	**2.441**	2.437	**2.535**	2.241	2.384
Macro-Financial	1.652	4.432	1.783	1.891	3.387	**2.465**	4.328	2.347	3.104
	Log-Predictive Score								
Behavioral	-1.974	**-1.613**	**-1.924**	-3.023	-2.457	-2.543	**-1.520**	-4.153	**-1.722**
Neoclassical	**-1.876**	-1.919	-1.936	-14.80	**-2.120**	**-1.301**	-2.668	**-0.879**	-2.652
Macro-Financial	-2.018	-1.640	-2.092	**-2.500**	-2.664	-2.569	-1.955	-3.161	-2.560
Panel C: Forecasting Horizon, $h = 12$									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
	RMSE								
Behavioral	**3.083**	**4.561**	**2.010**	**2.954**	1.802	3.599	4.542	**3.179**	**3.417**
Neoclassical	4.345	6.312	3.006	3.366	1.693	4.234	**3.952**	3.374	3.853
Macro-Financial	4.231	6.224	3.314	3.148	**1.527**	**3.520**	5.248	3.913	3.978
	Log-Predictive Score								
Behavioral	**-1.067**	**-3.260**	**-0.900**	-2.442	**-1.281**	**-0.423**	**-4.354**	-36.426	-5.441
Neoclassical	-1.311	-10.027	-1.806	-16.30	-1.597	-0.96	-8.593	-33.740	-12.72
Macro-Financial	-1.677	-5.269	-1.648	**-1.754**	-2.357	-2.187	-5.819	**-31.676**	**-3.992**

Table B.1. Testing the Number of Regimes

Testing the number of regimes. This table reports the results of a formal test for the number of regimes for each specification of regressors and across industries. Panel A shows the (log) marginal likelihoods and the corresponding Bayes factor in log-scale comparing the model with two vs. three regimes, and using stock valuation variables (related to the Behavioral hypothesis) as regressors. Panel B and C show the results obtained using industry-specific economic shocks (related to the Neoclassical hypothesis) and aggregate macroeconomic factors, respectively. The sample period is 1983:01-2014:12, monthly.

Panel A: Behavioral-related regressors									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
$K = 2$	-2219.83	-3397.67	-3058.13	-7303.63	-3126.65	-2701.69	-3231.81	-18208.61	-5962.29
$K = 3$	-2673.12	-4387.98	-3340.46	-8173.13	-3621.92	-3627.31	-4442.10	-20109.44	-7128.09
$\log_{10} \mathcal{B}_{2,3}$	65.47	436.14	105.19	341.73	209.87	594.56	498.41	870.66	493.35
Panel B: Neoclassical-related regressors									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
$K = 2$	-2419.78	-3358.81	-3066.74	-7652.38	-3195.01	-2660.80	-4185.99	-19269.53	-6928.04
$K = 3$	-2650.24	-4451.54	-3369.83	-8799.03	-3606.48	-3708.38	-5176.30	-21009.98	-7800.46
$\log_{10} \mathcal{B}_{2,3}$	100.20	475.10	131.78	498.54	178.90	455.47	430.57	756.72	379.31
Panel C: Aggregate Macro-Financial Factors									
	NoDur	Manuf	Enrgy	BusEq	Telcm	Shops	Hlth	Money	Other
$K = 2$	-2522.54	-3384.86	-3098.52	-7387.16	-3139.22	-2259.81	-3295.76	-18106.92	-5993.38
$K = 3$	-2695.21	-4374.86	-3508.54	-8070.36	-3755.05	-3804.27	-4394.50	-20415.20	-6911.64
$\log_{10} \mathcal{B}_{2,3}$	206.69	424.86	195.83	333.36	273.22	479.38	505.52	959.38	412.76

Table C.1. Summary of Inefficiency Factors

This table summarizes the inefficiency factors for the posterior values of the model parameters, estimated over the sample period 1983:01 - 2014:12. In order to assess inefficiencies we used as an example the model specification with the behavioral-related explanatory factors and $K = 2$, with $K = 1$ the state of no merger waves. The estimated inefficiency factors are based on the Bartlett kernel as in Newey and West (1987) with a bandwidth equal to 4% of the 10000 retained draws.

	Inefficiency Factor	
	Parameters	Mean
$\boldsymbol{\beta}$	10	6.1161
$\boldsymbol{\alpha}$	12	7.1011
$S^{1:T}$	500	8.3180

Table C.2. Summary of Convergence Diagnostics

This table summarizes the convergence results for the posterior values of the model parameters, estimated over the sample period 1983:01 - 2014:12. In order to assess inefficiencies we used as an example the model specification with the behavioral-related explanatory factors and $K = 2$, with $K = 1$ the state of no merger waves. 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 20% of the retained 10000 draws. The variances of the means are estimated with the Newey and West (1987) variance estimator using a bandwidth of 4% of the respective sample sizes.

Summary of Convergence Diagnostics			
	Parameters	5% Reject Rate	10% Reject Rate
$\boldsymbol{\beta}$	10	0.0431	0.0823
$\boldsymbol{\alpha}$	12	0.0322	0.0712
$S^{1:T}$	500	0.0000	0.0000

Figure 1. Directed Graph of Model Linkages

Model dependence structure. This figure shows the dependence structure which is implied by the Markov regime switching Poisson regression with time-varying transition probabilities as specified in Section 2 of the main text. z_t^m represents the collection of industry-specific and aggregate variables at time t for the industry $m = 1, \cdots, M$, S_t represents the industry-specific state of merger wave, $U_{ij,t}^m$ represents the utility index from the choice $j \in [0, 1]$ for firm $i = 1, \cdots, N$, $\epsilon_{ij,t}$ its unobservable individual attributes, and β_{S_t} the regime-dependent set of parameters.

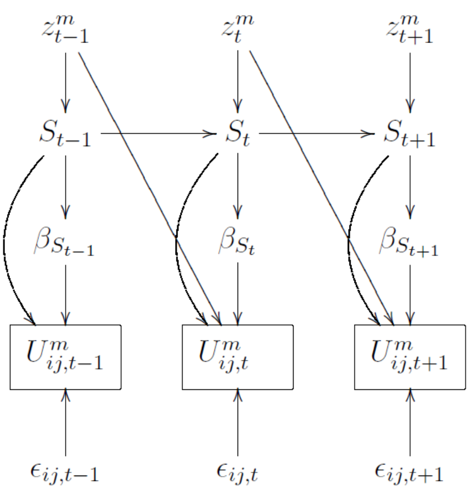

Figure 2. Time Series of Industry Merger Activity

Industry-specific merger activity over time. This figure reports the time series of merger activity across industries (grey bars), 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 1983:01 - 2014:12, monthly. We considered those deals with a value higher than 5 $ Mil (including net debt of the target), and exclude those identified as spinoffs, re-capitalizations, 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. Together with merger activity we show the inverse of the cross-sectional standard deviation of firms value-weighted book-to-market ratio (solid blue line). The left axis on each graph represent the number of deals, and the right axis show the scale for the regressor.

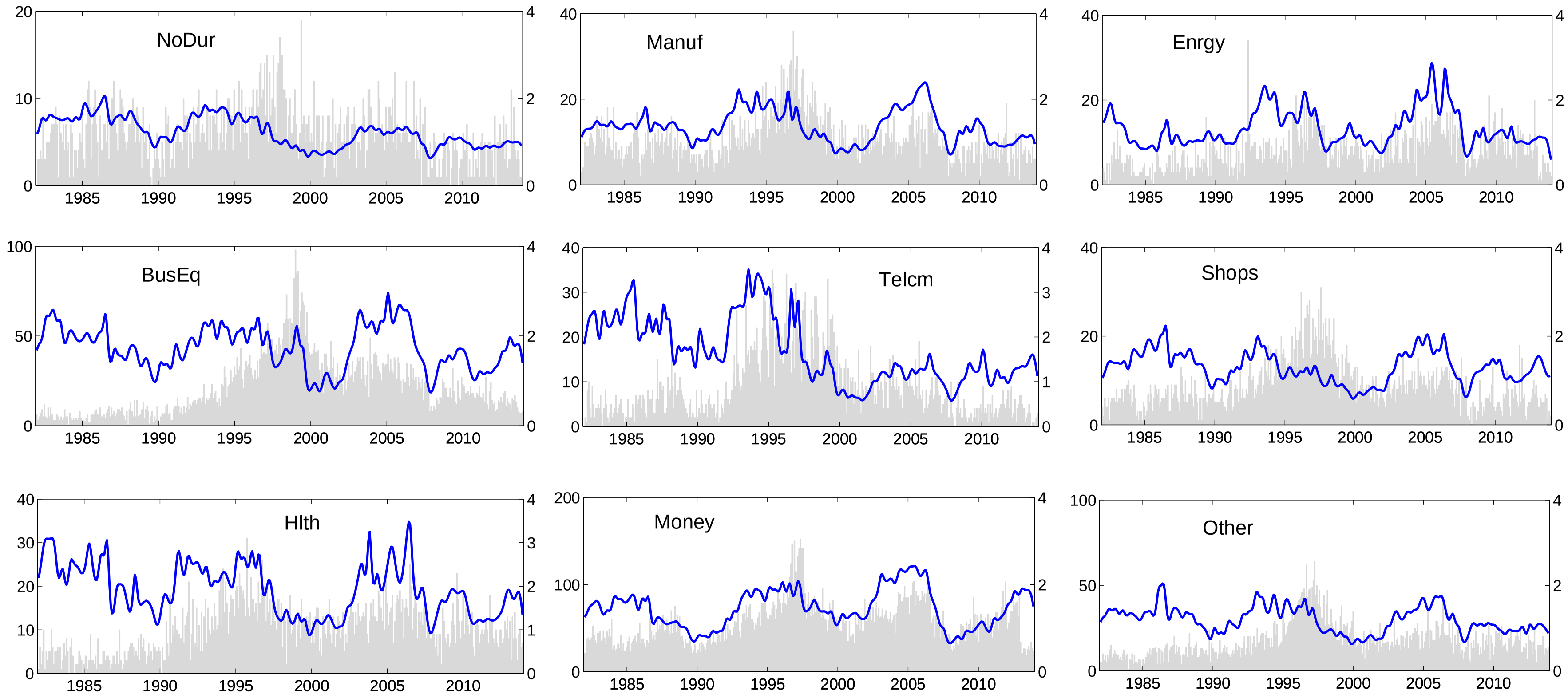

Figure 3. Effect of Behavioral-Related Factors on Merger Waves

The impact of industry-specific financial variables. This figure shows the impact of behavioral related N factors on the state of merger wave at the industry level as identified by the $N-$dimensional vector of coefficients $\alpha^z_{12} = \left(\alpha^z_{12,1}, \cdots, \alpha^z_{12,N}\right)$. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. As determinants we consider the industry-specific market-to-book ratio M/B and its cross-sectional standard deviation $sd(M/B)$, the industry-specific market returns (Ret), the aggregate market variance $(svar)$ and the market dividend-yield (dy). The sample period is 1983:01-2014:12, monthly. The box-plots report the median values for the changes (red line), the edges of the box are the $25th$ and $75th$ percentiles (blue box), and the whiskers extend to the 95% confidence intervals. Posterior estimates of the sensitivities conditional on the wave regimes are obtained from the Gibbs sampler detailed in Appendix A.

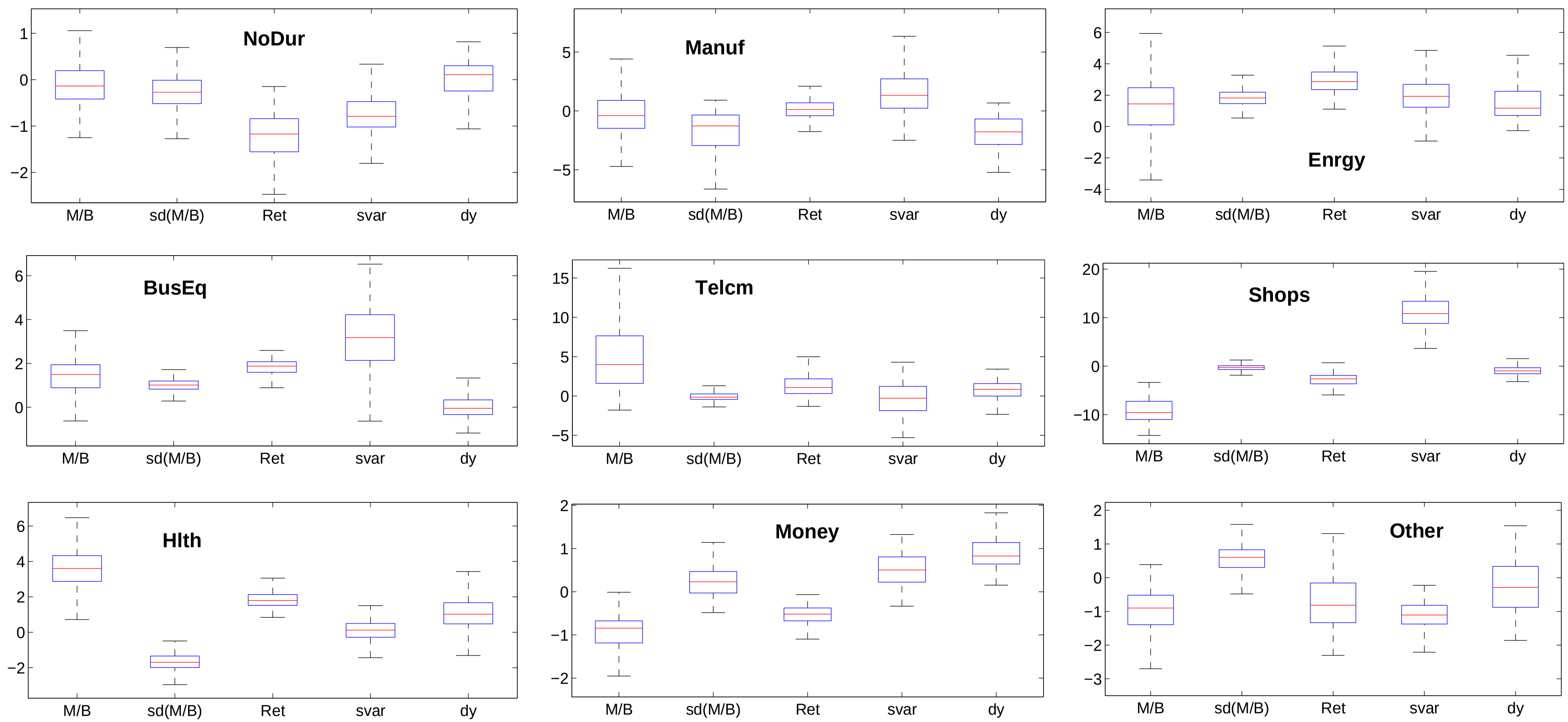

Figure 4. Effect of Neoclassical-Related Factors on Merger Waves

The impact of industry-specific economic shocks. This figure shows the impact of neoclassical related N factors on the state of merger wave at the industry level as identified by the $N-$dimensional vector of coefficients $\alpha_{12}^{z} = \left(\alpha_{12,1}^{z}, \cdots, \alpha_{12,N}^{z}\right)$. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. As determinants we consider median absolute changes of for cash-flow margin on sales (cash flow scaled by sales, *Marg*), asset turnover (sales divided by beginning-of-period assets, AT), return on assets (ROA), capital expenditures (scaled by beginning-of-period assets, $CAPX$), research and development (scaled by beginning-of-period assets, *R&D*), employee growth (*Emp*), sales growth (*Sale*), capital liquidity conditions (proxied by the spread between the average interest rate on commercial and industrial loans and the Federal Funds rate, *CL*), and a dummy variable identifying sector specific regulatory shocks where appropriate (*Reg*). The sample period is 1983:01-2014:12, monthly. The box-plots report the median values for the changes (red line), the edges of the box are the $25th$ and $75th$ percentiles (blue box), and the whiskers extend to the 95% confidence intervals. Posterior estimates of the sensitivities conditional on the wave regimes are obtained from the Gibbs sampler detailed in Appendix A.

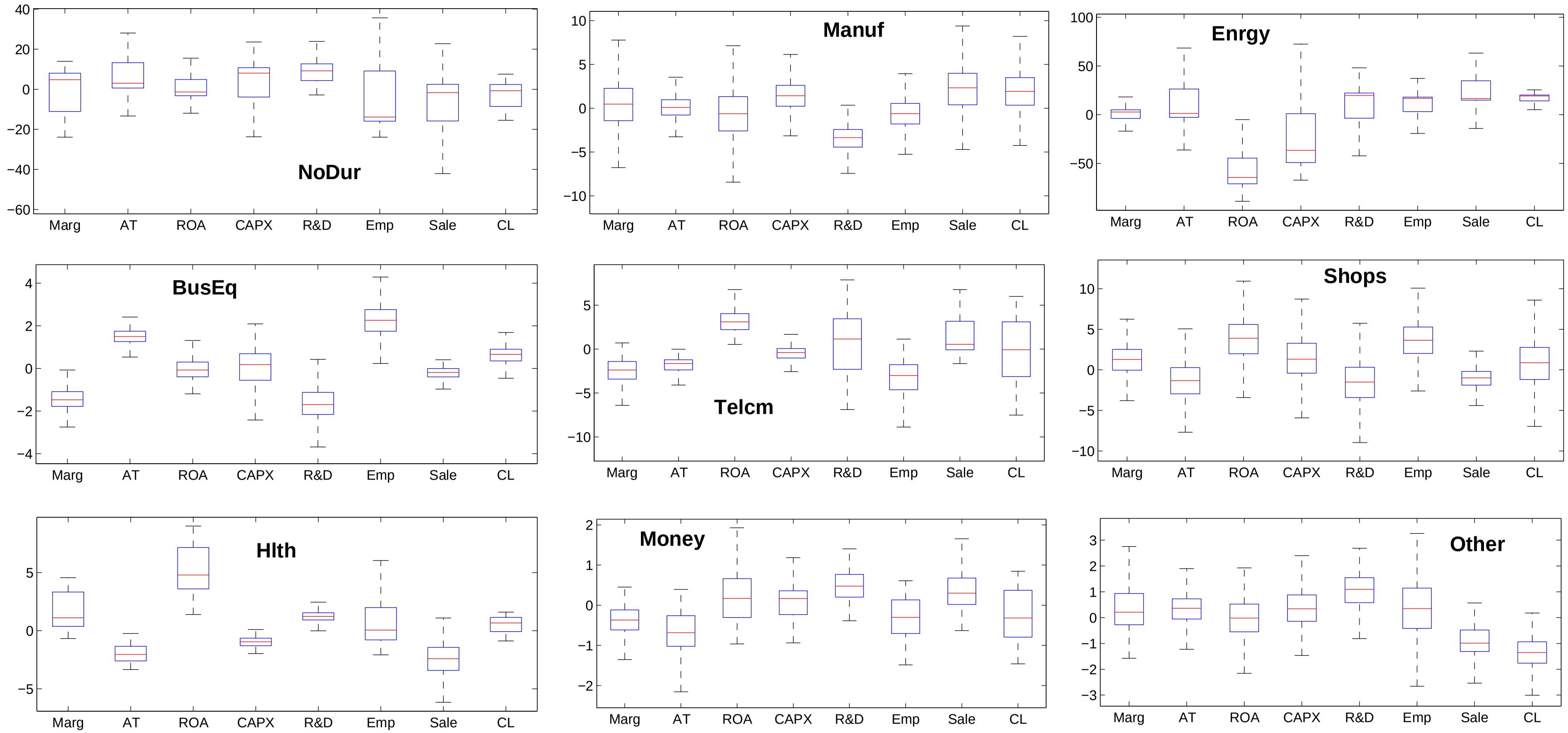

Figure 5. Effect of Macro-Financial Factors on Merger Waves

The impact of aggregate macro-financial factors. This figure shows the impact of N macro-financial factors on the state of merger wave at the industry level as identified by the $N-$dimensional vector of coefficients $\alpha_{12}^{z} = \left(\alpha_{12,1}^{z}, \cdots, \alpha_{12,N}^{z}\right)$. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. As determinants we consider the monthly compounded year-on-year industrial production growth (*IP*), the yield spread between the 10-year government bond and the 3-month T-Bill (*Term*), the yield spread between 20-year Baa and Aaa corporate bonds (*Credit*), the aggregate market-to-book ratio (*M/B*), the aggregate dividend yield (*dy*), the difference between the 1-month T-Bill rate and monthly inflation rate (*Rf*), and a dummy variable indicating a recession according to the NBER from the Peak through the Trough. The sample period is 1983:01-2014:12, monthly. The box-plots report the median values for the changes (red line), the edges of the box are the 25*th* and 75*th* percentiles (blue box), and the whiskers extend to the 95% confidence intervals. Posterior estimates of the sensitivities conditional on the wave regimes are obtained from the Gibbs sampler detailed in Appendix A.

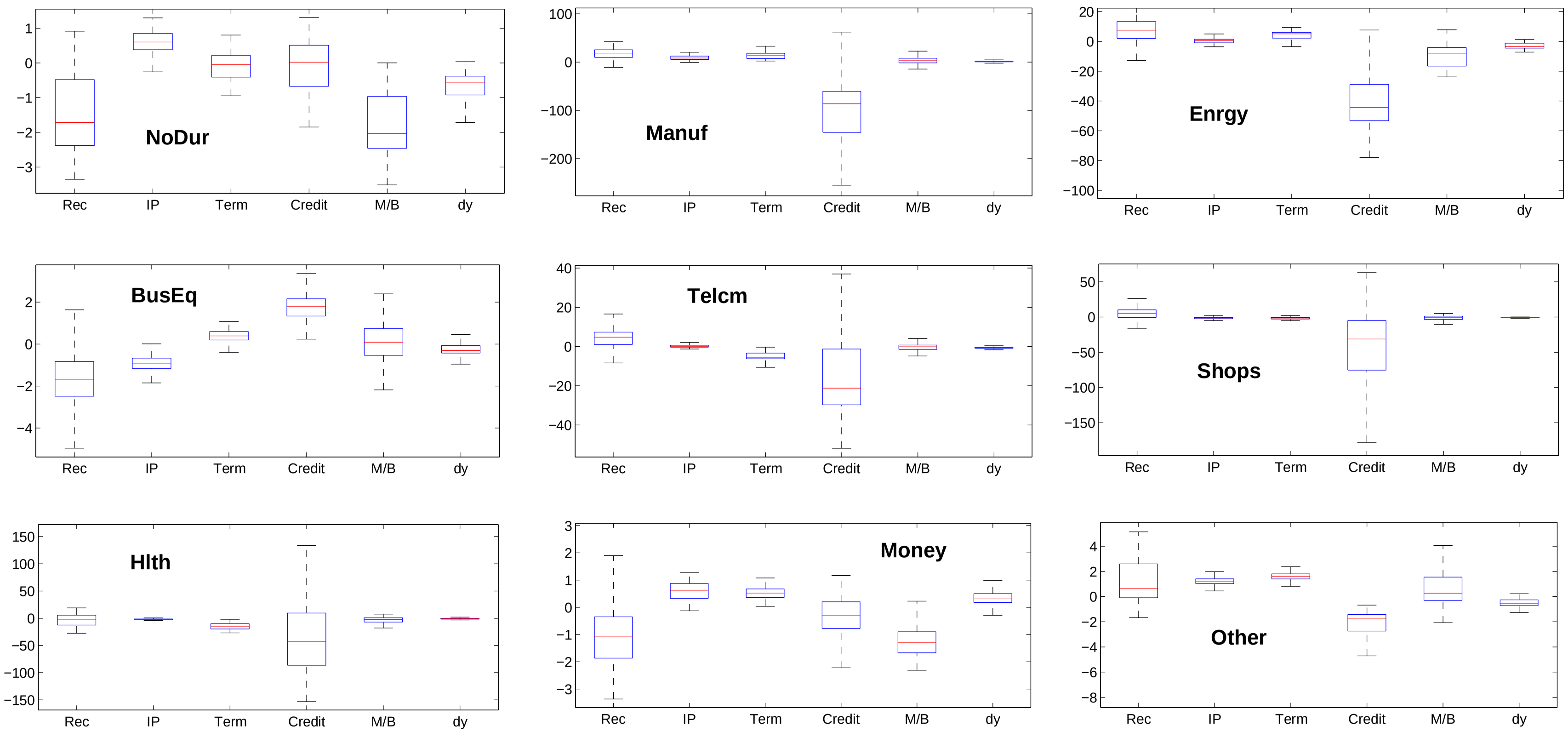

Figure 6. Posterior Probabilities of Industry Merger Waves Implied by the Behavioral Theory

Merger waves implied by industry-specific financial factors. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. This figure shows the average posterior probability obtained using the of regressors justified by a behavioral hypothesis on M&A deals flow. As determinants we consider the industry-specific market-to-book ratio M/B and its cross-sectional standard deviation $sd(M/B)$, the industry-specific market returns (Ret), the aggregate market variance ($svar$) and the market dividend-yield (dy). The grey area represents the posterior median probability of being in a state of merger wave, and the dashed blue line represents the historical number of M&A deals. The sample period is 1983:01-2014:12, monthly. Posterior average estimates of the probabilities are obtained from the Gibbs sampler detailed in Appendix A.

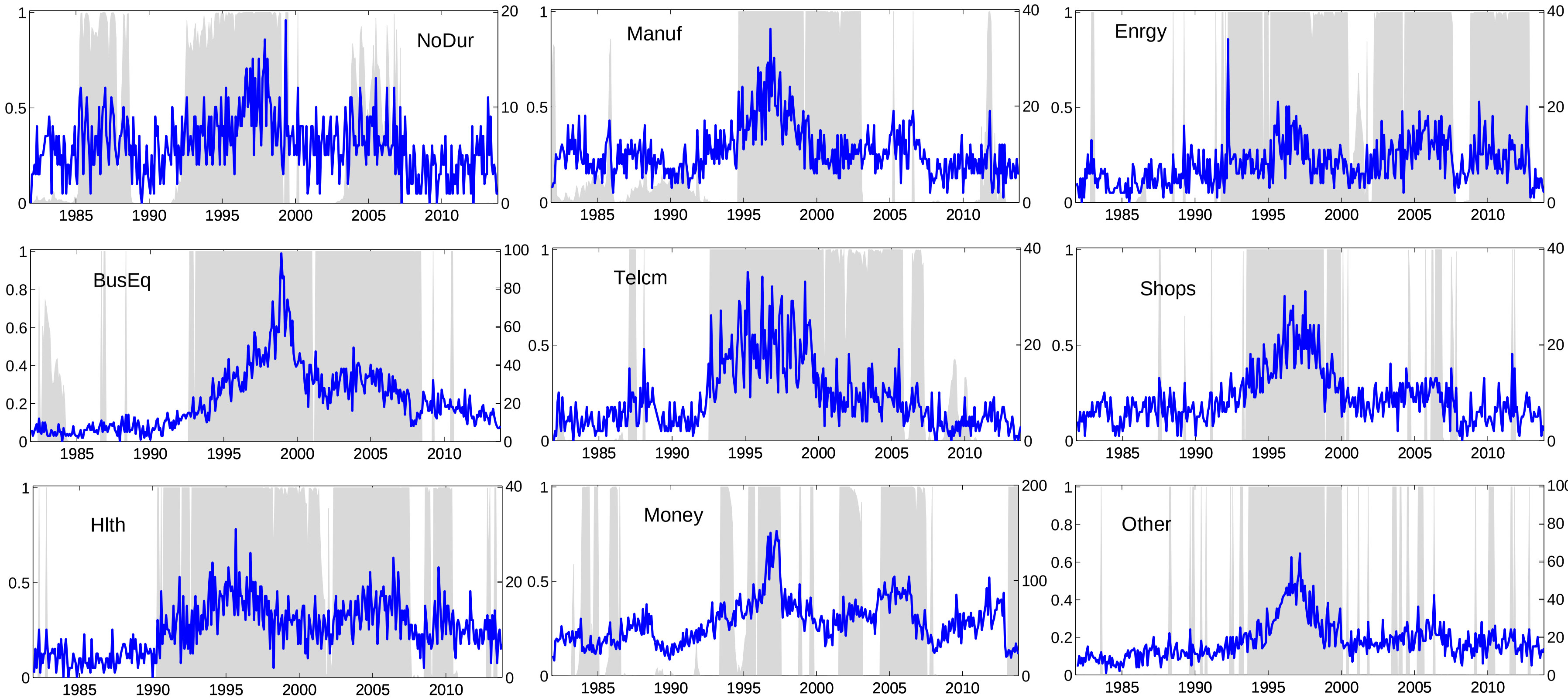

Figure 7. Posterior Probabilities of Industry Merger Waves Implied by the Neoclassical Theory

Merger waves implied by industry-specific economic shocks. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. This figure shows the average posterior probability obtained using the of regressors justified by a behavioral hypothesis on M&A deals flow. As determinants we consider median absolute changes of for cash-flow margin on sales (cash flow scaled by sales, *Marg*), asset turnover (sales divided by beginning-of-period assets, *AT*), return on assets (*ROA*), capital expenditures (scaled by beginning-of-period assets, *CAPX*), research and development (scaled by beginning-of-period assets, *R&D*), employee growth (*Emp*), sales growth (*Sale*), capital liquidity conditions (proxied by the spread between the average interest rate on commercial and industrial loans and the Federal Funds rate, *CL*), and a dummy variable identifying sector specific regulatory shocks where appropriate (*Reg*). The grey area represents the posterior median probability of being in a state of merger wave, and the dashed blue line represents the historical number of M&A deals. The sample period is 1983:01-2014:12, monthly. Posterior average estimates of the probabilities are obtained from the Gibbs sampler detailed in Appendix A.

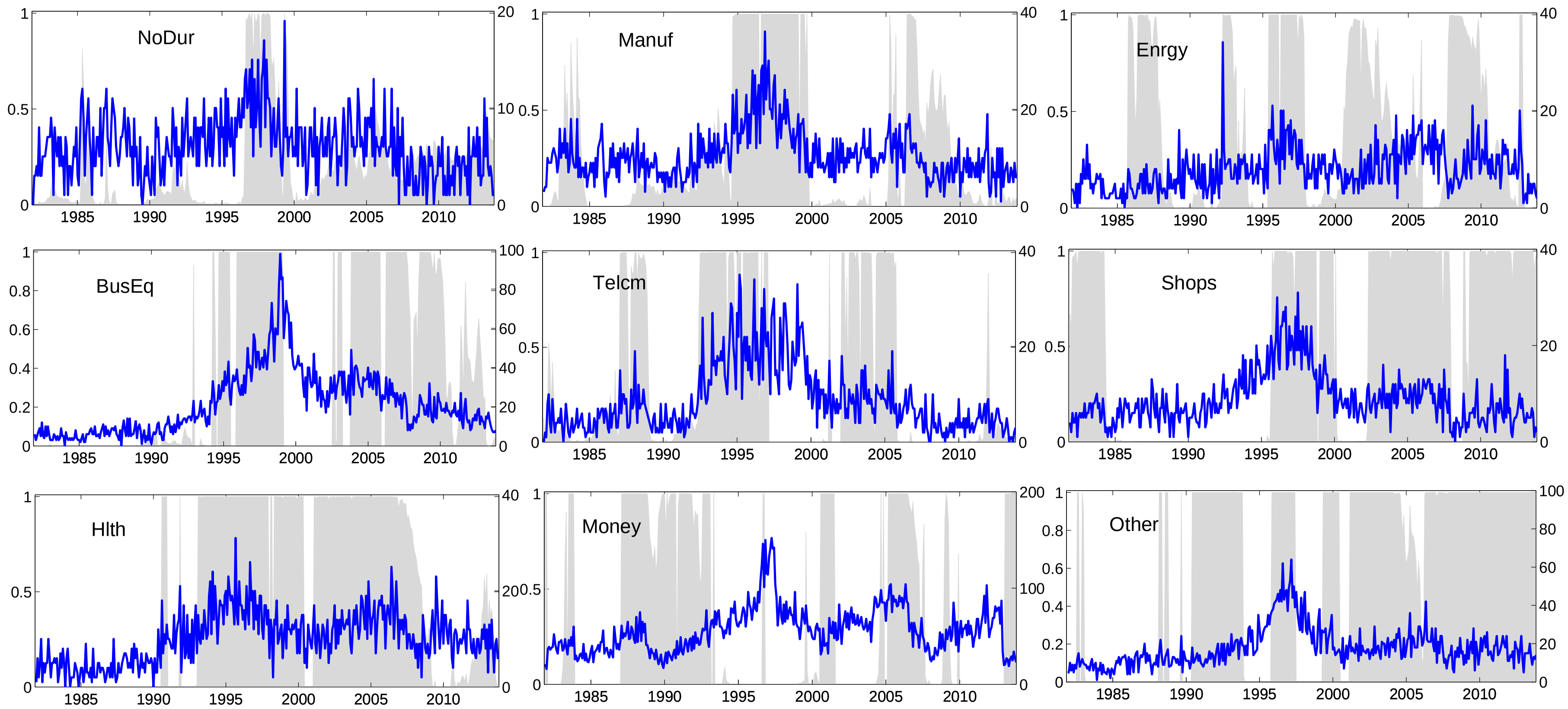

Figure 8. Posterior Probabilities of Industry Merger Waves Implied by Macro-Financial Factors

Merger waves implied by aggregate macro-financial factors. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. This figure shows the average posterior probability obtained using the of regressors justified by a behavioral hypothesis on M&A deals flow. As determinants we consider the monthly compounded year-on-year industrial production growth (IP), the yield spread between the 10-year government bond and the 3-month T-Bill ($Term$), the yield spread between 20-year Baa and Aaa corporate bonds ($Credit$), the aggregate market-to-book ratio (M/B), the aggregate dividend yield (dy), the difference between the 1-month T-Bill rate and monthly inflation rate (Rf), and a dummy variable indicating a recession according to the NBER from the peak through the trough. The grey area represents the posterior median probability of being in a state of merger wave, and the dashed blue line represents the historical number of M&A deals. The sample period is 1983:01-2014:12, monthly. Posterior average estimates of the probabilities are obtained from the Gibbs sampler detailed in Appendix A.

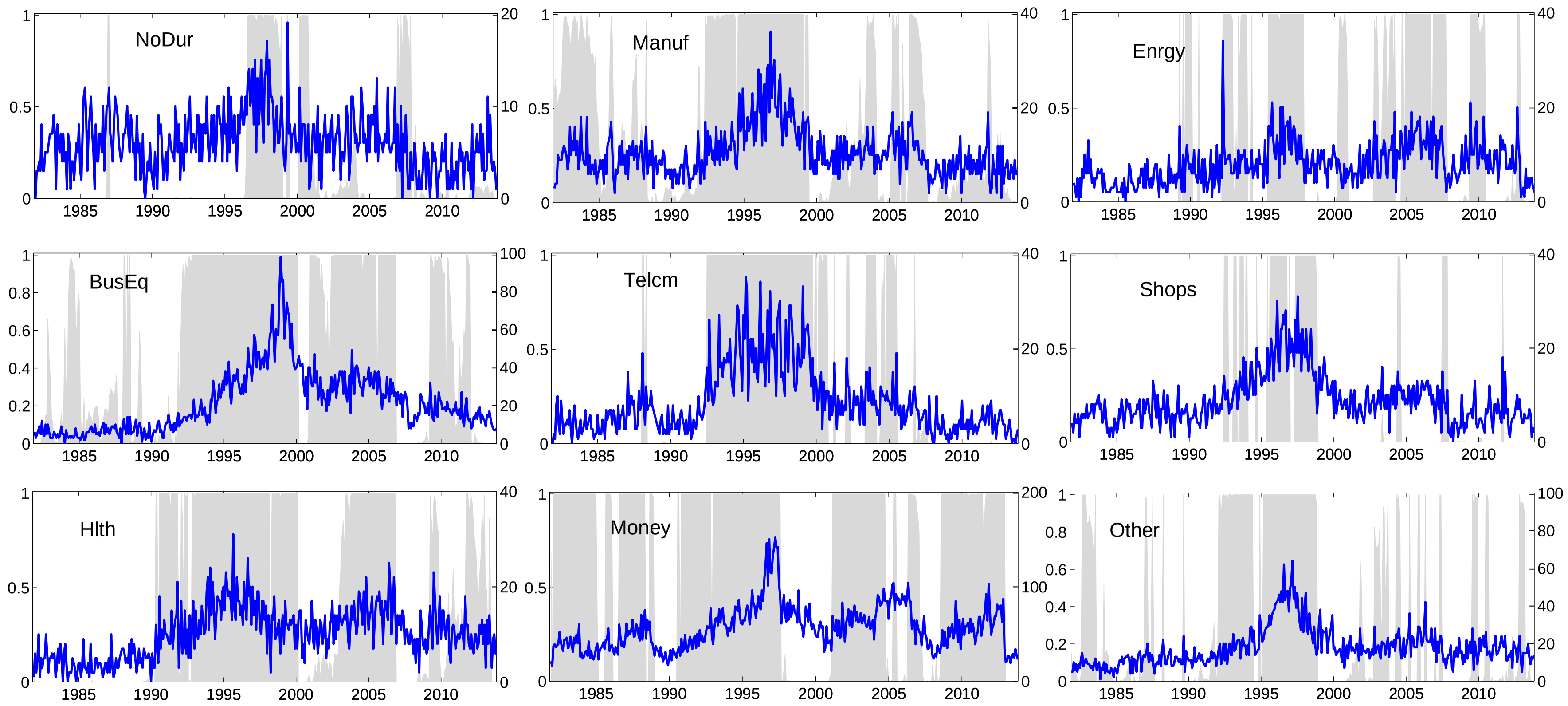

Figure 9. Industry Merger Activity and Model-Implied Intensity Rates

Model-implied intensity rates. This figure shows the actual number of M&A deals and the median model-implied intensity rates computed by considering different theory-based set of regressors. Deals (grey bars) 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 1983:01 - 2014:12, monthly. We considered those deals with a value higher than 5 $ Mil (including net debt of the target), and exclude those identified as spinoffs, re-capitalizations, 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 solid red line is the merger intensity rate implied by the model using the set of behavioral regressors; the dash-dot blue line represents the merger intensity rate explained by the aggregate macroeconomic variables, and the black-dashed line is the merger intensity rate implied by the set of industry-specific explanatory variables justified by the neoclassical hypothesis no M&A activity. Posterior median estimates are obtained from the Gibbs sampler detailed in Appendix A.

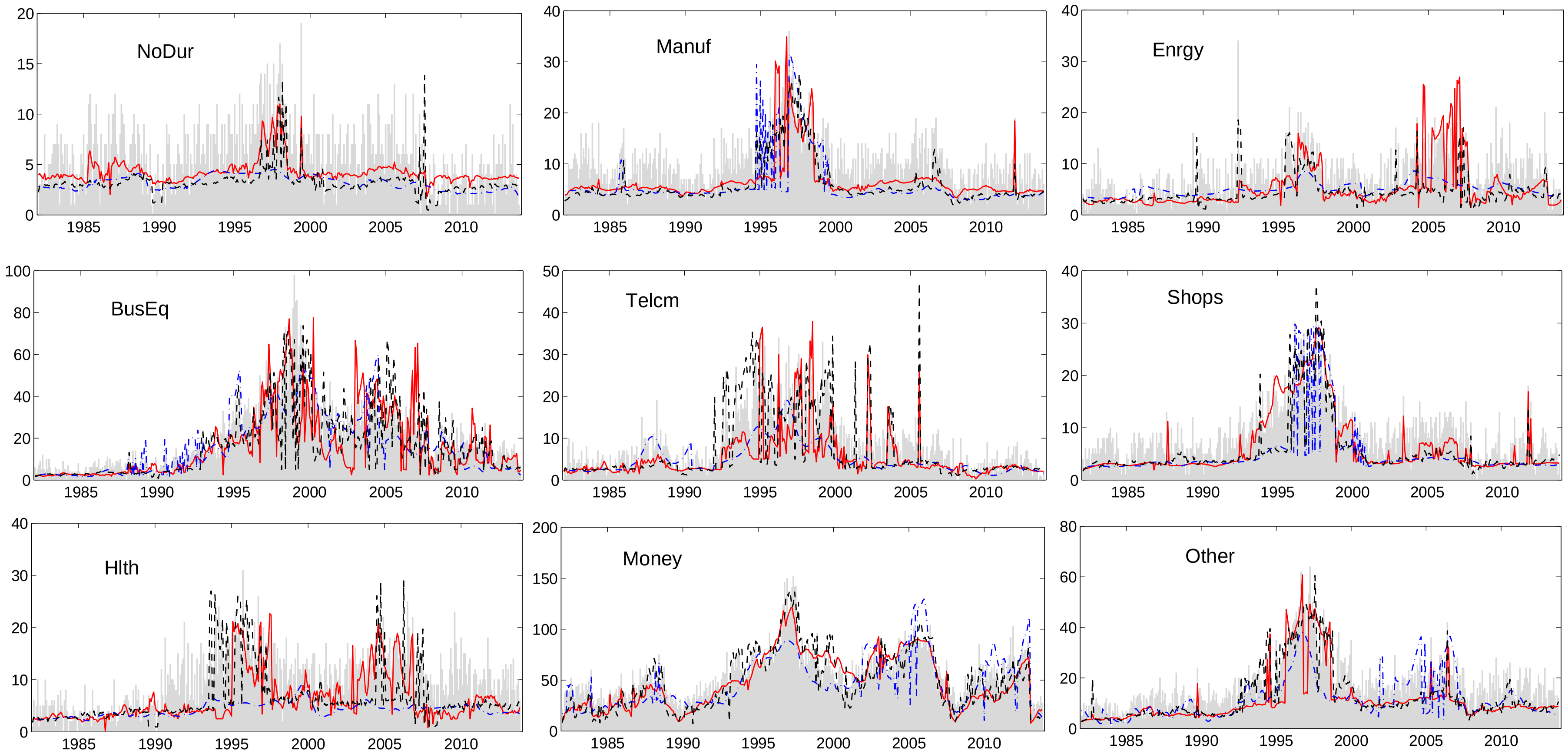

Figure 10. Coincidence Rates for Merger Waves Indicators

Cross-industry correlation of merger waves. This figure shows the in-sample pairwise coincidence rates of merger waves indicators between industries and for different theory-based set of regressors. We construct an indicator that takes value one if the model-implied waves indicate the same outcome for both industries (either wave or absence of wave) and zero otherwise. The heating maps report the probability of observing the same outcome. The sample period is 1983:01-2014:12, monthly. Estimates of the model-implied waves are based on posterior medians obtained from the Gibbs sampler detailed in Appendix A. Industry classification is based on the four-digit SIC codes according to the twelve-industry classification provided by Kenneth French. The left panel shows the heating map representing the correlation between industry-specific waves implied by neoclassical hypothesis determinants; the mid and right panels report the same cross-sectional correlation measures implied by the behavioral hypothesis-related regressors and macroeconomic factors, respectively.