



Institute for Strategy and Business

Economics

University of Zurich

Working Paper Series

ISSN 1660-1157

Working Paper No. 92

Are There Waves in Merger Activity After All?

Dennis L. Gärtner and Daniel Halbheer

August 2008

Are There Waves in Merger Activity After All?*

Dennis L. Gärtner[†] and Daniel Halbheer

August 2008

Abstract

This paper investigates the merger wave hypothesis for the US and the UK employing a Markov regime switching model. Using quarterly data covering the last thirty years, for the US, we identify the beginning of a merger wave in the mid 1990s but not the much-discussed 1980s merger wave. We argue that the latter finding can be ascribed to the refined methods of inference offered by the Gibbs sampling approach. As opposed to the US, mergers in the UK exhibit multiple waves, with activity surging in the early 1970s and the late 1980s.

Keywords: Merger Waves, Markov Regime Switching Regression Model, Gibbs Sampling

JEL Classification: G34, C32, C11, C15

*We are indebted to Dan Akerberg, Stefan Buehler, James Hamilton, Ariel Pakes, Patrick Rey, Armin Schmutzler, Enrique Schroth, Konrad Stahl, Ulrich Woitek, Lucy White, Michael Wolf, and seminar audiences at the University of Zurich, the 2004 EC2 conference on the econometrics of industrial organization, and the 2005 annual meeting of the Swiss Society of Economics and Statistics for most helpful comments and suggestions.

[†]Corresponding author. Address: University of Zurich, Socioeconomic Institute, Blümlisalpstr. 10, CH-8006 Zurich; tel.: +41 44 634 06 05; fax +41 44 634 49 07; e-mail: dennis.gaertner@soi.unizh.ch.

1 Introduction

There is broad consensus that mergers occur in waves. Since the seminal contribution by Nelson (1959), many studies have reported a wave-like pattern in merger activity, pointing out the merger waves of the mid 1980s and mid 1990s in the US in particular.¹ Guided by these observations, a vast empirical literature has sought to identify potential causes and triggers for merger waves.² This empirical strand has more recently been complemented by efforts to explain the phenomenon of merger waves in the theoretical literature.

While the general notion of mergers occurring in waves is practically undisputed, there is no clear consensus on how to operationalize the concept of a ‘merger wave’ in a time series context. The empirical literature has put forward three distinct approaches to modeling and identifying such waves. First, Golbe and White (1993) have sought to identify waves by fitting a sine curve to historic merger data. Second, merger series have been modeled by autoregressive processes capable of producing wave-like behavior (Shughart and Tollison, 1984; Clark et al., 1988; Chowdhury, 1993; Barkoulas et al., 2001). Third, and finally, merger series have been modeled by means of parameter-switching models where waves in activity are caused by discrete parameter switches (Town, 1992; Linn and Zhu, 1997).

This paper reexamines the case for detecting waves in merger activity in a time series context using more recent, consistent data and refined estimation techniques. Following Town (1992) and Linn and Zhu (1997), we employ a Markov regime switching model to describe the stochastic behavior of merger activity. We provide a thorough motivation for this approach, starting from Nelson’s (1959, p. 126) observation that aggregate merger series are characterized by “large bursts of activity separated by lengthy intervals of very low activity,” which we take to suggest the presence of two distinct unobserved states of merger activity, ‘high’ and ‘low’. By letting mean and variance of the autoregressive model be determined by realizations of the Markov process governing the evolution of the two states, waves are triggered by

¹Studies discussing the 1980s merger wave include Ravenscraft (1987), Golbe and White (1988, 1993), and Mitchell and Mulherin (1996). Mueller (1997), Andrade et al. (2001), and Harford (2005) describe the 1990s merger wave. An extensive review of earlier merger waves is provided in Scherer and Ross (1990, pp. 154–159).

²See e.g. Ravenscraft (1987), Shleifer and Vishny (1990), and Holmstrom and Kaplan (2001). Gugler et al. (2005) examine hypotheses that have been put forward as explanations of merger waves.

switches in the unobserved state. While this approach borrows from Town (1992) and Linn and Zhu (1997), we propose a slightly modified formal specification in which the autoregressive processes' inertia persists also *across* state switches, leading merger activity to react less abruptly to such switches. More importantly, we use new and consistent quarterly time series data covering merger activity both in the US and the UK, extending from 1973:IV through 2003:IV and from 1969:I through 2003:IV, respectively.³

In this paper, we challenge the notion of the much-discussed 1980s merger wave in the US. We argue that the discrepancy between our findings and previous econometric identifications of this wave is driven by a further distinguishing feature of our analysis: the use of more recent estimation techniques. To address the central issue of wave identification, we conduct inference on the regime indicator within a Bayesian framework employing Gibbs sampling techniques (Gelfand and Smith, 1990; Albert and Chib, 1993). In contrast, the aforementioned studies by Town (1992) and Linn and Zhu (1997) base wave identification on Maximum Likelihood techniques (Hamilton, 1989, 1993). In this latter approach, inference consists in first estimating the model's unknown parameters via Maximum Likelihood, and then conducting inference on the unobserved state conditional on the parameter estimates. Bayesian analysis, on the other hand, avoids this two-step procedure by treating both the model parameters and state variable as random variables and basing inference on states on a joint distribution of parameters and states rather than on a conditional distribution. This methodological difference can lead to quite different conclusions regarding the likely path of the unobserved regime indicator if parameter uncertainty is sufficiently high, as the uncertainty on parameter estimates does not feed into uncertainty on states when employing a two-step estimation procedure.

Our main results are as follows. *First*, we find that the US has witnessed only the beginning of a wave in merger activity, this wave starting in 1995:IV. This result is consistent with the observations in Mueller (1997), Holmstrom and Kaplan (2001), and Andrade et al. (2001), all of which report an upsurge in merger activity in the mid 1990s. However, our investigation of industry level data does not support the prominent notion that waves in aggregate merger activity represent the clustering

³Previous empirical studies examined the merger wave hypothesis using an assemblage of separate series differing in coverage and inclusion criteria. For a general discussion of available historical time series merger data and their limitations, see e.g. Golbe and White (1988).

of surges within one or a few industries (Mitchell and Mulherin, 1996; Mulherin and Boone, 2000; Andrade et al., 2001; Harford, 2005). *Second*, even when fitting the model only to the data prior to the estimated break date, we fail to identify the much discussed 1980s merger wave. To explain our difference in findings, we argue that if there is sufficient uncertainty surrounding the model’s parameters, then the two-step Maximum Likelihood estimation procedure can convey a deceptive degree of certainty about state inference. *Third*, the UK has witnessed two merger waves, the first starting in 1971:I and ending in 1973:IV and the second lasting from 1986:III to 1989:IV. The dating of these merger waves is close to the evidence reported in Hughes (1993, p. 16).

The remainder of the paper is organized as follows. Section 2 gives a short overview of theoretical explanations of merger waves. Section 3 briefly comments on the data employed for our empirical study, section 4 provides a thorough motivation the model used. In Section 5, we describe the inference problem and give a brief introduction to the Gibbs sampling approach. Section 6 presents the main results of our estimation both for the US and the UK series. Section 7 concludes.

2 Theoretical Explanations of Merger Waves: An Overview

This section reviews some of the the theoretical explanations for merger waves.⁴ The two main lines of reasoning are the following: First, contributions from the field of industrial organization typically identify reasons for strategic complementarities between (otherwise independent) merger decisions. By these complementarities, a merger between any two firms makes it more attractive for other firms to merge, so that a single merger may trigger a wave. Second, contributions from financial economics typically point out that merger waves may represent a simultaneous reaction of firms to exogenous changes in the economic environment.

More specifically, models put forward in the industrial organization literature typically explain merger waves as the sequentially rational equilibrium outcome of a game involving a series of merger decision. For instance, Nilsson and Sørsgard (1998)

⁴Interestingly, the literature focuses entirely on explanations for merger waves at the industry level. A notable exception is Toxvaerd (2008), whose model can be reinterpreted as a model of aggregate merger activity.

consider a two-stage model in which two pairs of firms sequentially decide whether or not to merge before engaging in product market competition. Firms contemplating the first merger take into account whether their decision encourages or discourages the second merger, and how this merger will affect their profits. In equilibrium, a merger wave is triggered if the incentives are such that the first merger profitably induces the second.

In a closely related paper, Faulí-Oller (2000) investigates the potential for strategic merger waves using a model in which two low-cost firms sequentially bid for high-cost firms before product market interaction. He finds that, for certain sets of parameter values (reflecting cost differences and market size), the subgame perfect equilibrium outcome is that the first bidder finds it profitable to trigger a takeover wave. The intuition for this result is that the first firm can exploit competition among the larger number of competitors, so the compensation required to induce the first merger is lower than for the subsequent merger. However, as the merger increases industry profits, the second firm finds it profitable to acquire the remaining low-cost firm although the compensation is higher.

In a more recent contribution, Qiu and Zhou (2007) model merger waves in a more elaborate game involving multiple rounds of negotiation before firms eventually interact in the product market. In each round, a randomly drawn firm gets to bid for a target firm of its choice.⁵ The authors show that, for some range of parameter values (again, reflecting market size and costs) the equilibrium outcome involves a series of mergers. The intuition for their result is again that individual mergers induce further strategic mergers that would otherwise not occur, allowing early merging firms to free-ride on the subsequent price increase.

Toxvaerd (2008) proposes an alternative explanation of merger waves employing a dynamic model where the value of being merged remains subject to random fluctuations in the exogenous economic environment (mirroring for instance technology or demand shocks). He assumes that a set of acquirers compete over time for scarce targets and that each acquirer has the option of buying one target, so the issue is whether to postpone an acquisition or raid the target immediately. It turns out that if the realization of the economic fundamental lies above some cut-off level at which the value from raiding is non-negative, the equilibrium outcome is that all potential

⁵The merger game ends either if a target rejects a proposal or if all drawn firms refrain from exerting the option to propose a merger.

acquiring firms raid the target firms simultaneously.⁶ Intuitively, it is the risk of being stranded without a firm to merge that triggers the preemptive merger wave.

While theory offers insight concerning horizontal merger activity, the literature is scant on the phenomenon of merger waves in vertically related industries. A notable exception is Avenel (2008), who considers a successive Bertrand oligopoly model where firms decide on (pairwise) vertical integration and the adoption of a cost-reducing technology before product market competition takes place. He shows that, depending on the fixed cost associated with the introduction of the new technology, any degree of vertical integration can arise in equilibrium. The intuition for this finding is that the integration and the adoption of the new technology lowers the profitability of integration of other firms; so if the fixed costs are low enough, vertical integration will go on until the industry is fully integrated.

In contrast to the industrial organization literature, where the focus is on how one merger is related to the other through the strategic interaction on the product market, the financial economics literature studies the incentives to merge due to exogenous changes in the stock market valuation of capital, or firms. Jovanovic and Rousseau (2002) treat mergers as reallocations of used capital, and provide a model of merger waves that is based on technological change and Tobin's q -theory.⁷ They assume that a firm's output is given as a function of its capital stock and its random state of technology, and they assume further that a firm has the opportunity to sell or buy capital in the market for used capital. The authors then calculate, conditional on the state of technology, the market value of one unit of capital within the firm and compare it to the price at which used capital trades. Depending on this value gap, firms either sell or buy bundled capital (that is, mergers take place), and do so increasingly if the interfirm dispersion capital values gets larger (mergers waves as "reallocation waves").⁸

The papers by Shleifer and Vishny (2003) and Rhodes-Kropf and Viswanathan (2004) depart from the neoclassical perspective and show that merger waves can

⁶Albeit in a dynamic model, the equilibrium outcome is similar to those in Faulí-Oller (2000) and Qiu and Zhou (2007), where an exogenous realization of low demand may trigger a merger wave.

⁷Faria (2008) interprets Tobin's q as a measure of managerial ability and considers mergers as a way to transfer organizational capital from firms with low managerial skills to firms with high managerial skills.

⁸Jovanovic and Rousseau (2008) build on the notion of reallocation waves and shows that exits and mergers should rise after a technological shock, and that exits should lead to mergers.

solely occur because of valuation issues. Shleifer and Vishny (2003) posit that financial markets are inefficient in the short-run and consider a model of acquisitions where the firms are valued incorrectly due to (market-wide) “investor sentiment”. Given market beliefs of firm values and the perceived synergy of the merger, they study the long-term wealth effects of an acquisition on the merging parties’ shareholders. They arrive at the conclusion that rising stock market valuations trigger a merger wave where relatively overvalued firms acquire their targets for stock (the choice of making the acquisition for stock instead of cash allows bidding shareholders to redistribute wealth away from the target shareholders to themselves).

Rhodes-Kropf and Viswanathan (2004) differs from the previous paper by explicitly taking into account the competition among bidders for an acquisition target. The authors consider a sequence of potential acquisitions and model the mergers as second-price auctions with equity bids (the crucial assumption is that the bidding firms do not know the value of the target due to both market-wide and firm-specific misvaluation). In this setting, a merger wave is defined as a sequence of two or more time periods in which the probability of a merger occurring is above the unconditional expected probability of a merger. One of the model’s main implications is that a large enough realization of the market-wide misvaluation can trigger a merger wave. Intuitively, the wave phenomenon results if the market is unable to fully self-correct and stays overvalued after a shock, which in turn triggers the merger in the consecutive period.

3 The Data

Our paper follows the majority of previous empirical studies, particularly Town (1992) and Linn and Zhu (1997), in using the number of transactions as the measure of historical merger activity.⁹ Specifically, we investigate the following series:

- (i) *The US merger series, covering 1973:I–2003:IV*. The time series data are taken from various issues of *Mergerstat Review*, a publication by FactSet Mergerstat LLC. The series reports publicly announced mergers, acquisitions and unit

⁹Other prominent measures of merger activity suggested in the literature are the dollar value of merger transactions (see e.g. Golbe and White, 1988, Scherer and Ross (1990) and the number of transactions relative to appropriate population totals (see e.g. Hughes, 1993, p. 16; Gugler et al., 2005).

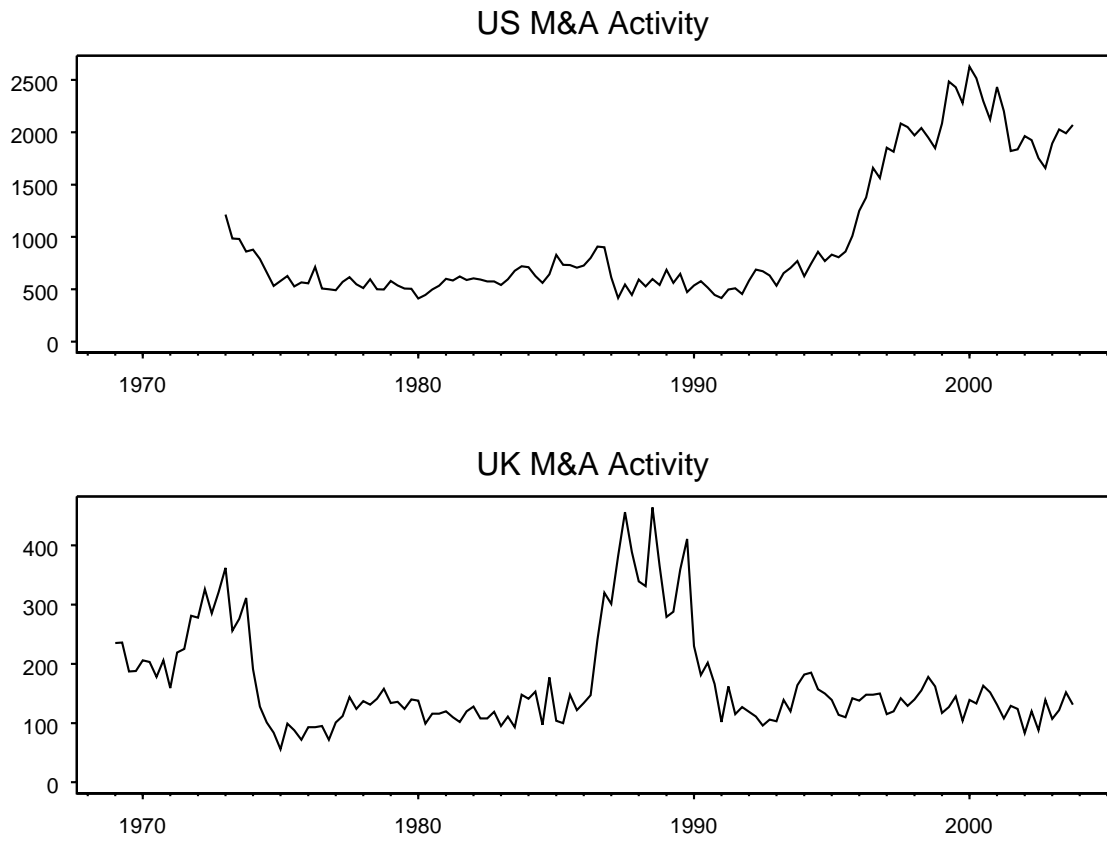


Figure 1: The Number of Merger Transactions, US Series, 1973:I–2003:IV (in Top Panel) and UK Series, 1969:I–2003:IV (Bottom Panel).

divestitures involving (i) at least one US company, (ii) a transaction volume exceeding \$1 million, and (iii) a purchase price exceeding 10% of the acquired company’s equity (i.e., an interest exceeding 10% of the acquired firm’s equity).

- (ii) *The UK merger series, covering 1969:I–2003:IV.* These data are published by the Office for National Statistics on a quarterly basis. The series consists of publicly announced mergers and acquisitions involving UK companies only. In contrast to the US data, there is no explicit cut-off bias relating to the value of the transaction, but the deal has to aim at gaining de jure control of the acquired company (i.e., a controlling interest exceeding 50% of the acquired firm’s equity).

Plots of these series are presented in Figure 1.

4 A Markov-Switching Model of Merger Waves

As outlined above, the literature has advanced the idea that mergers follow a wave pattern. We take this casual impression to suggest the presence of two distinct states of merger activity, high and low, as follows:¹⁰

Assumption 1. *Each period t is associated with an unobserved latent state variable $S_t \in \{1, 2\}$, where $S_t = 1$ implies that period t is a low-activity period and $S_t = 2$ denotes a high-activity period.*

The basic idea is then to let unobserved switches between states of high and low activity feed into observed merger activity—in a sense to be made precise shortly—so as to induce the alleged wave-like behavior. Hence, given Assumption 1, the remaining key questions concerning our description of mergers are: (1) What determines the unobservable state S_t in any period t , and (2) how exactly do the unobservable states feed into observed merger activity y_t ?

The general framework in which we deal with these questions is the Markov regime switching model originally proposed by Hamilton (1989). In a nutshell, this approach treats both the sequence of observations y_t and the sequence of states S_t as (interdependent) random variables, specifies a model which jointly generates the two sequences, and then estimates the model using the observed series y_t while treating the sequence of states as ‘missing data’. This framework offers several advantages over more traditional approaches to break-point analysis which typically rely on casual determination of candidate break-dates or *ad hoc* restrictions on the number of break dates (see e.g. Chow, 1960, 1984; Andrews, 1993): First and foremost, a major goal of our analysis is not only the estimation of regime-dependent structural model parameters, but dating the waves (i.e. conducting inference on the break dates themselves). This in turn requires modelling the probability law governing changes in regime rather than imposing particular break-dates *a priori*. Through the probability model, we can then let the data itself speak about the likely incidence of significant changes. Second, we would like to propose a unified structural process capable of describing various merger series (such as across countries or industries)

¹⁰We shall comment on the idea of using more than two states when discussing our estimation results further below. Let us just note for now that raising the number of attainable states invokes the usual trade-off between achieving a better fit to the data and overparameterizing the model. As a consequence, we suggest using the minimal number of states capable of producing the described behavior in mergers.

with apparently different frequency and timing of waves, which requires that wave dates be determined endogenously by the process.

More specifically, concerning the determination of states, we shall assume that states follow an independent first-order Markov process. Thus, in any period t , the probability of switching to a certain state in the next period $t + 1$ depends only on the state in period t . Specifically, we assume the following:

Assumption 2. *The unobserved state variable S_t follows a first-order Markov process with transition probabilities from any period t to period $t + 1$ given by*

$$\Pr(S_{t+1} = 1|S_t = 1) = p_{11} \quad \text{and} \quad \Pr(S_{t+1} = 2|S_t = 2) = p_{22}, \quad (1)$$

with $p_{11}, p_{22} \in [0, 1]$. In any period t , these transition probabilities are independent of past (log) merger realizations (y_t, y_{t-1}, \dots) .

It is important to note that ‘discrete merger waves’ as we understand and model them need not display a highly regular periodic pattern. Indeed, the first-order Markov specification implies that the process governing the states displays very little memory. This low-memory approach seems justified by the aforementioned literature giving little impression that the documented bursts of high activity display a highly regular periodic pattern.¹¹ Some structure is of course nonetheless implied by our Markov specification, such as the expected duration of a high state being $p_{22}/(1 - p_{22})$ and the expected duration of a low state being $p_{11}/(1 - p_{11})$, but these durations generally display a rather high variability. Furthermore, due to the first order Markov property, the *remaining* expected duration of a certain state is independent of how long the process has already been in that state, which again reflects the low-memory quality of the process.

Finally, note that the Markovian model encompasses the extreme possibility of a state being ‘absorbing’ in the sense that, once the process reaches a certain state, it remains in that state indefinitely (so that the regime switch is permanent rather than transitory). This is the case for the low-activity state if $p_{11} = 1$ and for the high-activity state if $p_{22} = 1$. Conversely, whenever this is not the case, so $p_{11}, p_{22} < 1$ and if in addition $p_{11} + p_{22} > 0$ (so there is no completely deterministic alternation

¹¹Although Golbe and White (1993) do report evidence of a sine wave pattern in US merger activity based on data up to 1989, by inspection of the plots in Figure 1 we strongly suspect that their model would no longer provide a very good fit to our more recent series.

between states), then the Markov chain turns out to be *ergodic* (see e.g. Hamilton, 1989). Then, a further key characteristic of the state switching model is given by the ergodic regime probabilities $\Pr(S_t = i)$, i.e. the unconditional probability of state $i \in \{1, 2\}$.¹² These can be shown to be given by $\Pr(S_t = 1) = (1 - p_{11}) / (2 - p_{11} - p_{22})$ and $\Pr(S_t = 2) = 1 - \Pr(S_t = 1)$.

Let us now turn to the second question concerning the exact form of the state's impact on merger activity. We start from the idea that merger activity follows some sort of mean reverting autoregressive process and augment this by assuming that both the mean and the variance of this process are time-varying and dependent on the states.¹³ Specifically, we make the following assumption:

Assumption 3. *Conditional on the sequence of unobserved states S_t , (log) mergers y_t follow the AR(k) process*

$$y_t - \mu_{S_t} = \sum_{i=1}^k \phi_i (y_{t-i} - \mu_{S_t}) + \varepsilon_t, \quad (2)$$

where (i) the ε_t are independently $N(0, \sigma_{S_t}^2)$ and independent of previous merger realizations $(y_{t-1}, y_{t-2}, \dots)$, (ii) $\mu_{S_t} \in \{\mu_1, \mu_2\}$ and $\sigma_{S_t} \in \{\sigma_1, \sigma_2\}$ are determined by the state in period t , (iii) $\mu_2 \geq \mu_1$, and (iv) the autoregressive coefficients $\phi_1, \phi_2, \dots, \phi_k$ are restricted so that the roots of the associated lag polynomial, $\phi(L) \equiv 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_k L^k$, lie outside the complex unit circle.

We let the idea that $S_t = 2$ entails higher activity impose the normalization $\mu_2 \geq \mu_1$, while σ_1 and σ_2 are left unrestricted (except for the obvious nonnegativity requirement). Furthermore, the familiar condition on the autoregressive parameters $\phi_1, \phi_2, \dots, \phi_k$ ensures that the process is in some sense mean reverting, where this mean however depends on the state sequence. Put somewhat differently, the condition ensures that the only source of non-stationarity is through switches in the

¹²A convenient way to think of the ergodic probabilities is in terms of the fraction of high and low states observed in an infinitely long realization of the Markov chain.

¹³A previous study by Shughart and Tollison (1984) reports little success in describing waves in merger activity as a standard autoregressive process with *constant* mean and variance, $y_t - \mu = \sum_{i=1}^k \phi_i (y_{t-i} - \mu) + \varepsilon_t$ with $\varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$, where the wave property would be reflected solely by some of the higher-order autoregressive coefficients ϕ_2, \dots, ϕ_k being nonzero. However, such a specification can produce only rather 'tame', linear wave-like oscillations, while we suspect that the large bursts of activity separated by long intervals of low activity identified in the aforementioned literature can only be reconciled with a nonlinear model such as ours.

regime (i.e. given a constant sequence of states, $S_t = i$ for all t and $i \in \{1, 2\}$, y_t follows a stationary process).¹⁴

Specification (2) differs in a small but important way from Hamilton's (1989) original specification,

$$y_t - \mu_{S_t} = \sum_{i=1}^k \phi_i (y_{t-i} - \mu_{S_{t-i}}) + \varepsilon_t, \quad (3)$$

which has been the workhorse model in the literature on mean and variance switching Markov models and also happens to be the model used by Town (1992) and Linn and Zhu (1997) to describe mergers in particular. The subtle but important difference is that specification (2) assumes 'sluggish' adjustments of the merger series to a state switch, whereas by specification (3), state switches cause an immediate full shift in activity. To see this, observe that in specification (2), what systematically affects today's deviation from the mean, $y_t - \mu_{S_t}$, is a weighted sum of past deviations from the *current* mean, $y_{t-i} - \mu_{S_t}$, whereas in specification (3), today's deviation is determined by past deviations from the *contemporaneous* mean, $y_{t-i} - \mu_{S_{t-i}}$. Thus, the two models imply rather different dynamic consequences of a shift in regime. This is most effectively illustrated by setting $k = 1$ in both (2) and (3) above and considering a permanent shift from state 1 to state 2 between dates t and $t + 1$. According to specification (3), the switch to state 2 at date t raises the value of any subsequent y_{t+j} ($j \geq 0$) by $\mu_2 - \mu_1$ over its respective value if no state-switch had occurred. In model (2), on the other hand, the impact of the state switch at t only raises subsequent y_{t+j} by $(1 - \phi_1^j)(\mu_2 - \mu_1) \leq \mu_2 - \mu_1$ for any $j \geq 0$.¹⁵ Hence, model (3) suggests that the merger series immediately jumps toward the new mean after a state switch, whereas model (2) describes a more gradual, 'sluggish' gravitation toward the new mean. Note however that the difference of the state switch's impact between the models disappears as j rises, so that the models differ most markedly during the adjustment period.

¹⁴The literature has also proposed non-mean reverting processes such as random walks to describe merger activity (see e.g. Chowdhury, 1993). Even though standard tests reject the unit root hypothesis for our UK merger series, this is indeed not the case for the US series. However, it is a well understood fact that in general, unit-root tests have very little power over Markov-switching alternatives (see e.g. Nelson et al., 2001), so that such tests do not invalidate our proposed model. Furthermore, if we perform the unit root tests using only US data prior to 1995:III (which amounts to discarding little more than a quarter of the data), the unit root hypothesis is clearly rejected. We take this as evidence favoring our Markov-switching model over the random walk hypothesis.

¹⁵Recall that in an AR(1) model, $|\phi_1| < 1$ by the restriction on the autoregressive coefficients in Assumption 3.

We favor specification (2) over (3) for two somewhat interrelated reasons. First, casual inspection of real merger data suggests that the transition to a significantly higher (or lower) level of merger activity is indeed sluggish rather than immediate. Second, perhaps contrary to other common applications of mean switching models, there seems to be no intuitive reason to suggest that the merger process does not display the same amount of inertia when switching to a high or low activity state as within a given state. Indeed, if for instance we suspect the sluggishness in merger series to be a consequence of the fact that real world mergers may take considerable time to process (due to preparation, approval, etc.), thereby causing sluggish adjustment to any unobserved structural shocks, then this sluggishness should persist also when the economy moves to a generally higher or lower level of activity (i.e. when it is hit by a ‘large’ shock).¹⁶ For these reasons, we shall employ specification (2) for the remainder of our analysis.¹⁷

As a final remark, we should point out that more generally, mean and variance switching is not the only way in which high and low activity states may be thought to affect mergers. For instance, an alternative specification might have states impact only the growth rate rather than the mean level of the merger series.¹⁸ However, we view the mean-switching specification as closest in spirit to the wave notion developed in the literature. What may nonetheless seem somewhat extreme at first sight is that our mean switching model appears to posit that waves always have the same magnitude (or, stated differently, the described ‘bursts of activity’ always have the same magnitude).¹⁹ However, we would like to argue that empirically, a major task in identifying waves is being able to tell actual waves from smaller ‘ripples’, and a straightforward way to accomplish this is to posit that waves always have a certain height. We will return to this point in our discussion of the 80s merger wave in Section 6.3.

¹⁶This argument can be formalized by noting that specification (2) can be interpreted as a standard $AR(k)$ model where the state only affects the distribution of the error term. This can be seen by rewriting model (2) as $y_t = \sum_{i=1}^k \phi_i y_{t-i} + \tilde{\varepsilon}_t$, $\tilde{\varepsilon}_t \sim N(\tilde{\mu}_{S_t}, \sigma_{S_t}^2)$, where $\tilde{\mu}_{S_t} = (1 - \sum_{i=1}^k \phi_i) \mu_{S_t}$.

¹⁷A nice technical side-effect of using specification (2) is that inference on states does not involve an approximation (see e.g. Kim and Nelson, 1999, pp. 68–70).

¹⁸For instance, Owen (2004) pursues such an idea for UK mergers by fitting a three-state mean switching Markov model to the differenced level data $\Delta y_t \equiv y_t - y_{t-1}$.

¹⁹Note that in our model with sluggish adjustment to the mean, this is only strictly true in expectation for waves having the same duration. If the process is sluggish, then the shorter a wave, the lower its peak.

5 Estimation Techniques

We estimate the model parameters and the path of the latent Markov switching regime indicator within a Bayesian framework employing Markov chain Monte Carlo simulation methods. Letting $\boldsymbol{\beta} \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, \phi_1, \phi_2, \dots, \phi_k, p, q)$ denote the model's parameters, letting $\mathcal{Y}_T = (y_1, y_2, \dots, y_T)$ denote the data observed, and letting $\mathcal{S}_T \equiv (S_1, S_2, \dots, S_T)$ denote the unobserved sequence of states, Bayesian inference in our model takes the form of using the data \mathcal{Y}_T and the model specified in Section 4 to map a given prior distribution of parameters, $p(\boldsymbol{\beta})$, into a joint posterior distribution of states and parameters, $p(\mathcal{S}_T, \boldsymbol{\beta} | \mathcal{Y}_T)$.

Rather than investigating $p(\mathcal{S}_T, \boldsymbol{\beta} | \mathcal{Y}_T)$ analytically, Markov chain Monte Carlo methods provide a simple way of simulating draws from this distribution. We use a particular form of these methods, the Gibbs sampling technique, which is an iterative scheme based on simulating successive draws from the conditional posterior distributions of the state vector \mathcal{S}_T and the appropriately partitioned parameter vector $\boldsymbol{\beta}$:

- (i) $p(\mathcal{S}_T | \boldsymbol{\beta}, \mathcal{Y}_T)$
- (ii) $p(\mathbf{p} | \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\phi}, \mathcal{S}_T, \mathcal{Y}_T)$
- (iii) $p(\boldsymbol{\mu} | \boldsymbol{\sigma}, \boldsymbol{\phi}, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T)$
- (iv) $p(\boldsymbol{\sigma} | \boldsymbol{\mu}, \boldsymbol{\phi}, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T)$
- (v) $p(\boldsymbol{\phi} | \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T)$,

where $\boldsymbol{\mu} \equiv (\mu_1, \mu_2)$, $\boldsymbol{\sigma} \equiv (\sigma_1, \sigma_2)$, $\boldsymbol{\phi} \equiv (\phi_1, \phi_2, \dots, \phi_k)$, and $\mathbf{p} \equiv (p_{11}, p_{22})$.²⁰

In contrast to the *full* posterior $p(\mathcal{S}_T, \boldsymbol{\beta} | \mathcal{Y}_T)$, each of the *marginal* posterior distributions (i)–(v) can be handled analytically. Simulated draws from (i)–(v) are thus easily obtained, and the Gibbs sampler provides a way of iterating on such draws to simulate draws from the full posterior $p(\mathcal{S}_T, \boldsymbol{\beta} | \mathcal{Y}_T)$.²¹

²⁰For a general introduction to Gibbs sampling, readers are referred to Geman and Geman (1984) and Gelfand and Smith (1990). A textbook treatment of the method can also be found in Kim and Nelson (1999).

²¹The derivation of the marginal posteriors is standard (cf. Albert and Chib, 1993; Kim and Nelson, 1999). In particular, the derivation of the marginal posteriors on states in step (i) uses Hamilton's (1989) filter—an iterative scheme for calculating the marginal probabilities $p(S_t | y_1, \dots, y_t, \boldsymbol{\beta})$. These filtered probabilities in turn may be used to compactly formulate the likelihood function in terms of a telescoping product in lieu of brute-force summation over all 2^T possible state sequences. See Kim and Nelson (1999) for details.

Given the tool to generate representative sets of draws from $p(\mathcal{S}_T, \boldsymbol{\beta}|\mathcal{Y}_T)$, properties of this distribution such as individual parameters' marginal distributions and moments are easily characterized by use of their sample equivalents. To address the central issue of wave identification, we will be particularly interested in $\Pr(S_t = 2|\mathcal{Y}_T)$, the posterior probability of being in a high merger activity state at any date t . These probabilities are obtained by averaging across the simulated paths for the states, each simulated while cycling through the above posterior distributions.

Basing state inference on $\Pr(S_t = 2|\mathcal{Y}_T)$, while natural in the Bayesian framework of Gibbs Sampling, elegantly overcomes a potential pitfall to classical inference. In a classical setting, inference on states is obtained through a two-step procedure by first obtaining a Maximum Likelihood parameter estimate $\hat{\boldsymbol{\beta}}$, and then calculating $\Pr(S_t = 2|\mathcal{Y}_T, \hat{\boldsymbol{\beta}})$, the probability of $S_t = 2$ in any period t under the assumption that $\hat{\boldsymbol{\beta}}$ corresponds to the *true* parameter values.²² From a Bayesian perspective, the derived inference on states is thus to be read as contingent on the econometrician having full confidence in his parameter estimate $\hat{\boldsymbol{\beta}}$. But only rarely will this correspond to the econometrician's true confidence in $\hat{\boldsymbol{\beta}}$. Moreover, alternative conceivable values of $\boldsymbol{\beta}$ will typically lead to different values of $\Pr(S_t = 2|\mathcal{Y}_T, \boldsymbol{\beta})$, so that uncertainty about $\boldsymbol{\beta}$ will feed into uncertainty on states. As a result, basing state inference on $\Pr(S_t = 2|\mathcal{Y}_T, \hat{\boldsymbol{\beta}})$ rather than on $\Pr(S_t = 2|\mathcal{Y}_T)$ can convey a false degree of certainty about states by neglecting uncertainty about parameters.

To make this important point more transparent, note that $\Pr(S_t = 2|\mathcal{Y}_T)$ and $\Pr(S_t = 2|\mathcal{Y}_T, \hat{\boldsymbol{\beta}})$ are related by

$$\Pr(S_t = 2|\mathcal{Y}_T) = \int_{\boldsymbol{\beta}} \Pr(S_t = 2|\mathcal{Y}_T; \boldsymbol{\beta})p(\boldsymbol{\beta}|\mathcal{Y}_T)d\boldsymbol{\beta}, \quad (4)$$

where $p(\boldsymbol{\beta}|\mathcal{Y}_T)$ denotes the posterior density of the parameter vector. By (4), our Bayesian inference based on $\Pr(S_t = 2|\mathcal{Y}_T)$ can be interpreted as considering $\Pr(S_t = 2|\mathcal{Y}_T; \boldsymbol{\beta})$ for *any* conceivable parameter constellation $\boldsymbol{\beta}$ —that is, for the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ in particular, but also for any other conceivable $\boldsymbol{\beta}$ —, weighting it with the respective posterior probability of $\boldsymbol{\beta}$, and adding up across $\boldsymbol{\beta}$ to produce $\Pr(S_t = 2|\mathcal{Y}_T)$. It is straightforward to see from (4) that if posterior parameter uncertainty is sufficiently high and conditional inference on S_t is sufficiently sensitive

²²For details on standard maximum likelihood methods, refer to Hamilton (1989, 1993).

to β , then $\Pr(S_t = 2|\mathcal{Y}_T, \hat{\beta})$ can differ substantially from $\Pr(S_t = 2|\mathcal{Y}_T)$.²³ We will provide a strong illustration of this difference below, where we shall in fact argue that this methodological point is likely to have played a substantial role in previous identifications of a 1980s merger wave.

6 Estimation Results

This section reports the results of fitting our lagged mean and variance switching model as given by Assumptions 1 through 3 to US and UK log merger series.²⁴ In a first step, Section 6.1 analyzes the entire quarterly US merger series from 1973 through 2003. Its main finding, the identification of a wave beginning in the mid 1990s, is augmented by a look at industry level data in Section 6.2. In a second step, to investigate the lack of evidence for a discrete 1980s merger wave in more detail, Section 6.3 reestimates the model using only data up to 1995. Section 6.4 argues that in so doing, the estimation techniques described in Section 5 play a decisive role. Finally, Section 6.5 offers a shift of focus to the UK by fitting the model to the quarterly UK merger series spanning 1969 through 2003.

As outlined, inference is conducted in a Bayesian context using the Gibbs sampling method to derive posterior distributions of the model parameters and to assess the sequence of unobserved states. In all cases, Gibbs sampling involved 20,000 iterations, where a burn-in sequence consisting of the first 1,000 draws was discarded prior to any inference. Output of the sampler was closely monitored to ensure proper convergence of the filter.

6.1 The Tidal Wave of the Mid 1990s: US Mergers, Full Series 1973:I through 2003:IV

First, we consider the full series of US mergers over the entire available time span from 1973:I through 2003:IV, as presented in Figure 1. Preliminary estimation of

²³See also Gärtner (2007) for a simple stylized example illustrating this point.

²⁴The main reason for using log rather than level merger data is that all series considered are nonnegative by construction. Strictly speaking, the model as defined by (1) and (2) therefore provides no valid description of the level series due to its capability of producing negative observations. However, all subsequent inference was nonetheless also conducted after fitting the level merger data to the model. Qualitative results, specifically concerning inference on states, differ only little from those obtained for the log merger series. Any remaining noteworthy differences are explicitly pointed out in the subsequent discussion.

	priors		posteriors				ML
	mean	std	mean	median	std	95%-band	MLE
μ_1	6.7	1000.00	6.386	6.390	0.083	(6.204, 6.538)	6.3959
μ_2	6.7	1000.00	7.687	7.677	0.103	(7.510, 7.924)	7.6653
σ_1	0.2	0.28	0.137	0.136	0.008	(0.123, 0.153)	0.1320
σ_2	0.2	0.28	0.096	0.095	0.010	(0.079, 0.118)	0.0834
ϕ_1	0.0	1.00	0.683	0.683	0.098	(0.488, 0.874)	0.6578
ϕ_2	0.0	1.00	0.063	0.063	0.113	(-0.159, 0.287)	0.0538
ϕ_3	0.0	1.00	-0.114	-0.114	0.112	(-0.334, 0.103)	-0.1136
ϕ_4	0.0	1.00	0.146	0.145	0.085	(-0.021, 0.313)	0.1386
p_{11}	0.9	0.21	0.987	0.991	0.012	(0.955, 1.000)	0.9886
p_{22}	0.9	0.21	0.997	1.000	0.010	(0.969, 1.000)	1.0000

Note: Result of 20,000 Gibbs-Sampling iterations, iterations 1,000 through 20,000 used for inference. 95%-band refers to 95% posterior probability bands.

Table 1: Estimation Results for US Merger Activity, 1973:I through 2003:IV.

the model with various lag lengths k suggested setting $k = 4$.²⁵

For all model parameters, Table 1 gives summary statistics both on the priors used and on the marginal posterior distributions obtained. The priors on all parameters were chosen to be relatively uninformative within the class of admissible conjugate priors (which are: normal distributions for the parameters μ_{S_t} and ϕ_i , inverted gamma distributions for $\sigma_{S_t}^2$, and beta distributions for p_{11} and p_{22}). To give an impression of the posterior parameter distributions beyond the mere summary statistics, Figure 2 plots histograms representing the estimated marginal posterior distributions of the parameters. Despite our focus on Bayesian inference, Table 1 also gives maximum-likelihood point estimates of all parameters in the last column.

The first feature obvious from Table 1 is that the data indeed leads to significant updates in the priors on all parameters, as shown by direct comparison of the standard deviations of priors and posteriors. Furthermore, Figure 2 reveals that the corresponding marginal posterior distributions are single peaked and well behaved. As may have been expected from a glance at the original data, comparing the posteriors for μ_1 and μ_2 reveals that mean log merger activity in the high state 2 significantly exceeds that in the low state 1.

²⁵More specifically, starting from $k = 10$, the number of lags was reduced until the 90% posterior probability band on the highest-order autocorrelating coefficient did not contain zero. Estimation results for various lag lengths also showed, however, that state-inference is little affected by the choice of lag length.

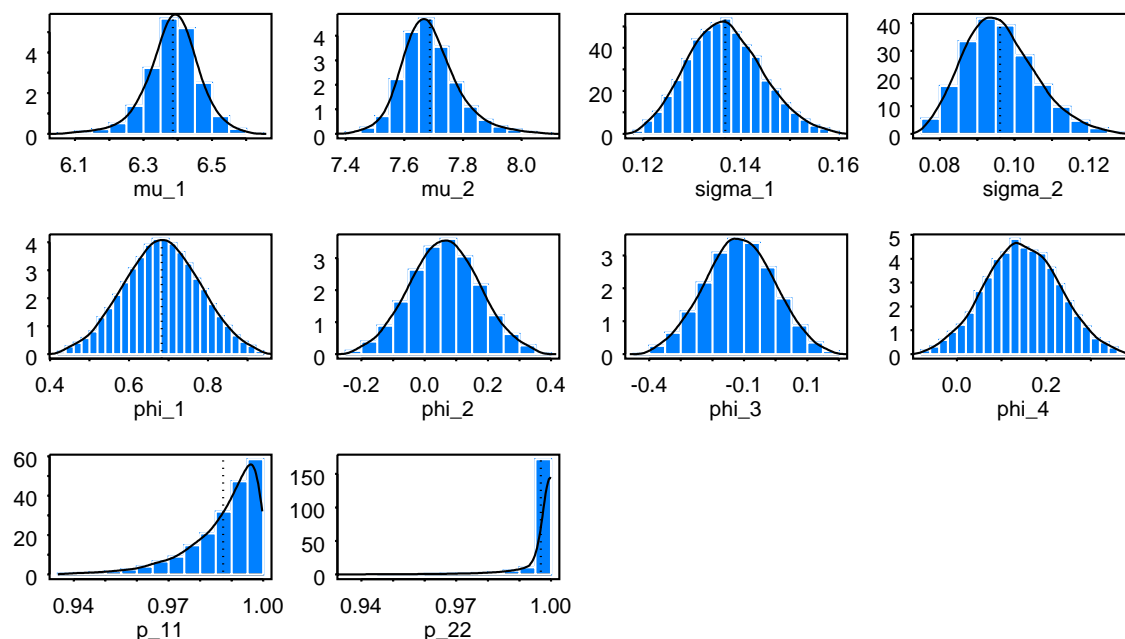


Figure 2: Marginal Posterior Distributions of Parameters for US Log Merger Series 1973:I–2003:IV.

Although not literally interpretable as state-contingent means of the untransformed series due to the nonlinear log transformation, the median values of $\exp(\mu_1)$ and $\exp(\mu_2)$ suggest that in level terms, merger activity in the low and high activity state are in the region of 596 and 2,160 mergers per quarter, respectively. Furthermore, log mergers seem to be significantly less volatile in the high activity state, as shown by a comparison of the posteriors on σ_1 and σ_2 . Again however, caution is called for when drawing conclusions concerning the volatility of the *level* merger series, as the log transformation compresses differences at higher absolute levels of activity. Indeed, fitting the model to the level merger data suggests that in level terms, mergers are significantly more volatile in the high-activity state. Next, estimates on the autoregressive coefficients ϕ_i show that mergers display a considerable degree of inertia also *within* states. Moreover, mean and median of the largest root across samples both figure at 0.84, suggesting a significant degree of inertia in the merger series. Finally, the posteriors on the transition probabilities p_{11} and p_{22} let us conclude that switches in regime are rather unlikely. Specifically, the median expected duration of a low activity state, $p_{11}/(1 - p_{11})$ is approximately 27 years (mean expected duration is heavily influenced by the skewness in the posterior on

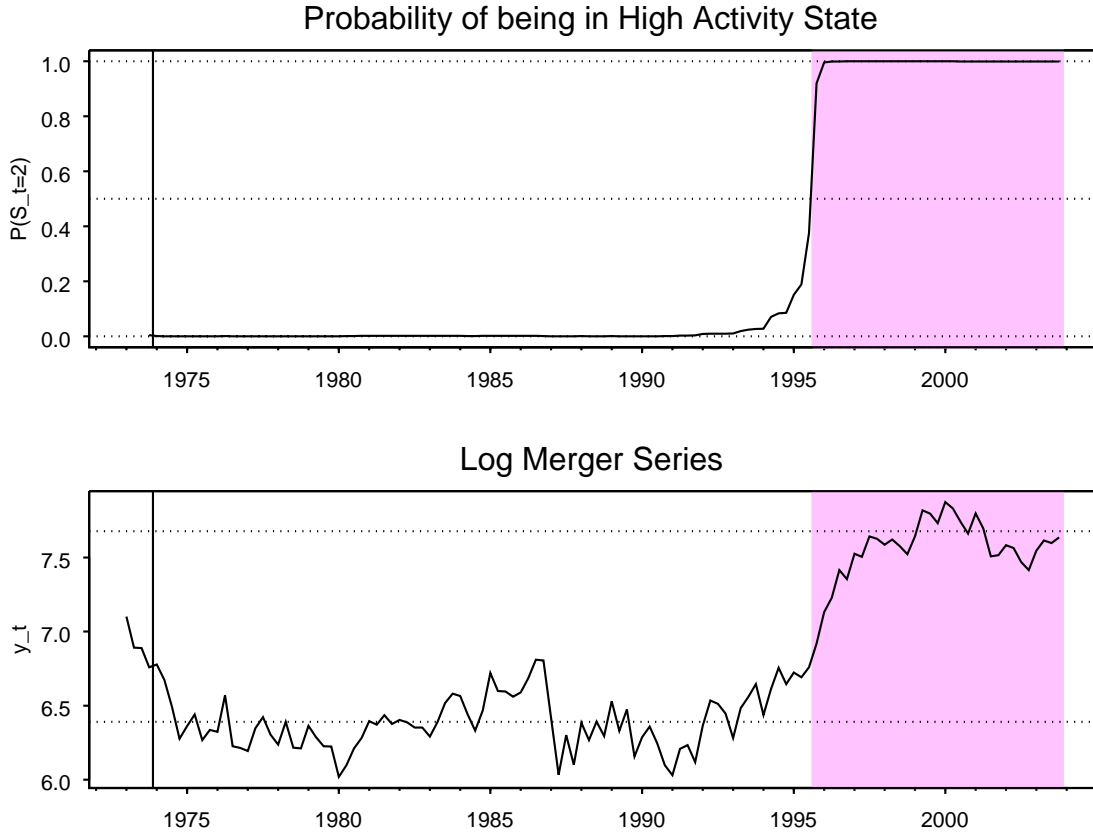


Figure 3: Estimated Probability of Being in State of High Merger Activity, US Log Merger Series 1973:I–2003:IV (in Top Panel, Bottom Panel displays Log Merger Series).

p_{22} and lies around 133 years). Conversely, estimates on p_{22} show that the high activity state seems to be essentially absorbing, so that a regime of high merger activity is unlikely ever to be left again—this result being driven, of course, by the fact that there does not seem to have been a single reversion from high to low activity in the series so far. Finally, for the ergodic probability of being in a high state (unconditional on the data), $\Pr(S_t = 2) = (1 - p_{22}) / (2 - p_{11} - p_{22})$, the mean posterior is 86.5%, whereas the median is 99.7%.

With these results on the model’s parameters in mind, let us now return to our main objective, the identification of waves in mergers. Figure 3 plots the probability of being in a high state in any period t given the observed US merger data, $\Pr(S_t = 2 | \mathcal{Y}_T)$. This probability plot is shown in the top panel of Figure 3, whereas the bottom panel reproduces the underlying log merger series (along with the posterior median of μ_1 and μ_2 as dashed horizontal lines) for convenience. To highlight the

most likely state for any period t , periods for which $\Pr(S_t = 2|\mathcal{Y}_T) > 0.5$ are shaded. However, any interpretation of this ‘best guess’ should take account of the underlying value of $\Pr(S_t = 2|\mathcal{Y}_T)$ as a straightforward measure of confidence in this guess: The closer $\Pr(S_t = 2|\mathcal{Y}_T)$ to 0.5, the more uncertainty surrounds the best guess.

Figure 3 shows that our estimation produces strikingly clear inference concerning the unobserved state. First, we find compelling evidence that over the entire period between 1973 and 2003, US merger activity has in fact experienced only a single regime switch, that switch being from low to high activity. Cast into the wave terminology, this suggests that since 1973, the US has so far witnessed only the beginning of a single ‘tidal wave’ in merger activity. Second, while this observation alone may come as no major surprise given a glance at the log merger plot, the clear-cut jump in the probability plot also allows us to date the beginning of this tidal wave rather precisely. Specifically, the assessed probability of a high state of merger activity jumps from 0.189 in 1995:II and 0.375 in 1995:III to a value of 0.920 in 1995:IV. We may thus conclude that this wave in US merger activity is very likely to have been triggered between the third and fourth quarter of 1995.²⁶

Rounding up, we should stress three points relevant to the interpretation of these results. First, as pointed out in Section 5, using $\Pr(S_t = 2|\mathcal{Y}_T)$ —rather than $\Pr(S_t = 2|\mathcal{Y}_T, \hat{\beta})$ for a point estimate of the parameters $\hat{\beta}$ —means taking account of uncertainty about the model’s structural parameters for the inference on states. It is all the more noteworthy that Figure 3 conveys an appreciably clear message concerning the likely sequence of states. Second, although we shall more thoroughly investigate the failure to identify the hump in merger activity around the mid 1980s as a discrete wave in a moment in Section 6.3, let us note for now that this result is even more clear-cut when fitting the state-switching model to the level rather than the log-merger data. Intuitively, this is due to the simple fact that the log transformation exaggerates the mid 1980s hump in merger activity relative to the level data. Finally, we should point out our estimation procedure’s weakness in producing inference on states at the very beginning of the series: For technical reasons, inference on states actually begins only in 1974:I (rather than in 1973:I),

²⁶Specifically, the probability of the US having witnessed a *single* state switch from low to high between 1995:III and 1995:IV (rather than at any other date) can be calculated at 54.2%, which is contrasted by the probability of a corresponding single switch one quarter earlier (i.e. between 1995:II and 1995:III) of 18.5%, of 3.7% two quarters earlier, and 7.5% one quarter later.

and inference on states in these early periods is generally somewhat sensitive to starting values used.²⁷

6.2 Reflections of the US 1990s Wave at the Industry Level

As a prominent explanation of the clustering of merger activity in time, the literature has advanced the idea that surges in aggregate merger activity represent firms' optimal responses to industry-level shocks.²⁸ According to this hypothesis, waves in merger activity at the aggregate level will be the result of temporary surges in merger activity in one or a few industries. Concerning our above findings, this naturally raises the question of whether the marked increase in US merger activity in the mid 1990s was dominated by any specific industries. While the available data do not permit us to estimate our model at the industry level, casual investigation of annual industry level data suggests that the mid 1990s wave is hardly attributable to one or a few industries alone.

To this end, of the 50 industries identified by the *Mergerstat Review*, Figure 4 plots *annual* merger data for those eleven US industries with the strongest merger activity between 1990 and 2003, where industries were ranked according to overall activity (in terms of numbers of mergers) for that period. While Figure 4 shows that industries were certainly not uniformly hit by a wave in 1995, it is nevertheless apparent that the resulting aggregate wave is anything but the result of a single industry level burst. For instance, the largest industry level share in overall annual merger activity was in Computer Software, Supplies and Services at its pronounced peak in 2000, with a share of 26%. Although significant, such shares still make it

²⁷To understand the first point, note that our estimation of an AR(k) model takes the first k observations of y_t as exogenously given, which is why inference on states begins in 1974:I. Concerning the second point, inferring states by means of the Hamilton filter requires specifying the initial exogenous probability of being in a high activity state, $\Pr(S_0 = 2|\beta; y_{-k+1}, \dots, y_0)$. All inference shown was produced using an uninformative but nonetheless somewhat arbitrary initial probability of 0.5. However, simulations with alternative initial probabilities were run to check the results for robustness. Specifically, for the full US merger series considered here, results turn out to be very robust despite the relatively high level of merger activity at the very beginning of the series. Intuitively, this stems from the fact that, even though the data suggest that the US may have been in a high level of merger activity immediately prior to 1973, the strong downward trend at the beginning of the sampling period clearly leads us to conclude that the US economy must nevertheless have already found itself in a low state of merger activity at the time our inference on states begins. More specifically, for an initial prior probability of 0.5, the posterior probability of the US economy having starting out in a low-activity state is 0.997.

²⁸See for instance Mitchell and Mulherin (1996), Mulherin and Boone (2000), Andrade et al. (2001), and Harford (2005).

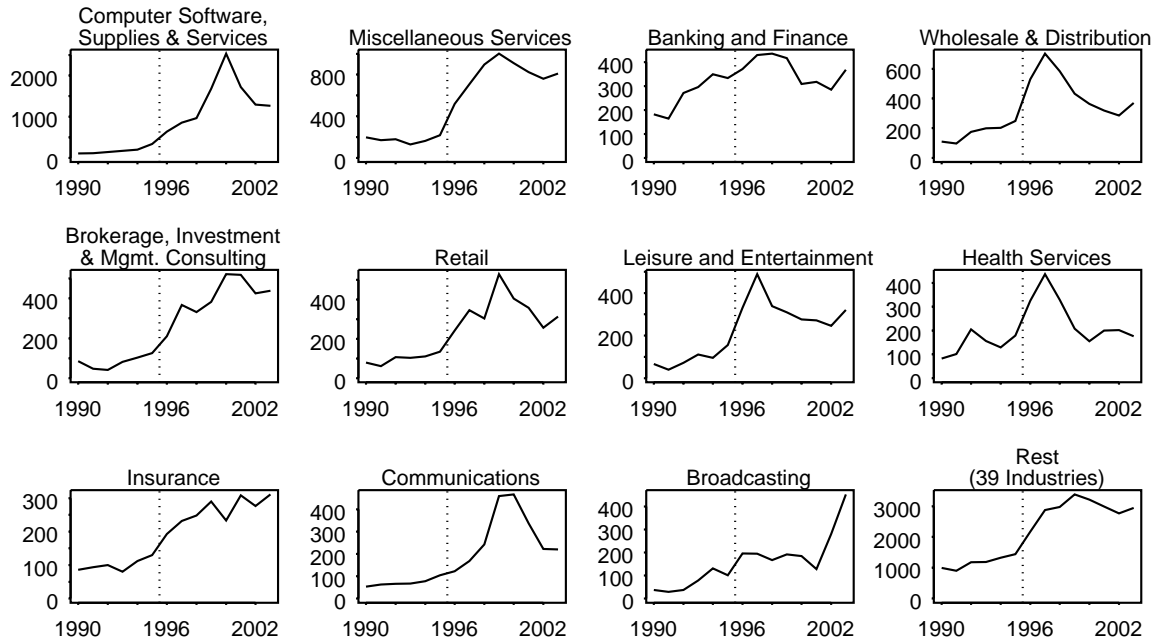


Figure 4: Annual US Industry Level Merger Activity, 1990 through 2003 (Source: *Mergerstat Review*).

impossible to explain the pronounced jump in aggregate mergers—from around 600 per quarter before the wave to over 2,000 thereafter—as caused only by a small subset of industries. Furthermore, as the bottom right plot in Figure 4 shows, even after removing the 11 most active industries (which account for 52% to 66% of annual merger activity between 1990 and 2003), residual merger activity still gives a strong impression of a mid 1990s merger wave.

6.3 What ever happened to the 1980s Merger Wave? US Mergers 1973:I through 1995:III

One of the most striking findings in Section 6.1 has been our failure to identify a regime-switch around the time of the much-discussed 1980s merger wave. Indeed, a simple look at the log merger series depicted in Figure 3 may indeed raise questions about there being something less pronounced—but nonetheless ‘wave-like’—about the visible hump in merger activity in the mid 1980s.

As pointed out previously, an arguably rather strong assumption implicit in our two-state model is that waves always come in similar sizes (where ‘size’ refers to a wave’s peak height). Along these lines, a perfectly valid reservation with our result

	priors		posteriors				ML
	mean	std	mean	median	std	95%-band	MLE
μ_1	6.4	1000.00	5.804	6.274	1.179	(2.467, 6.451)	6.3129
μ_2	6.4	1000.00	6.604	6.588	0.206	(6.323, 7.037)	6.7007
σ_1	0.1	0.14	0.125	0.128	0.036	(0.053, 0.200)	0.1222
σ_2	0.1	0.14	0.105	0.105	0.030	(0.055, 0.161)	0.0753
ϕ_1	0.0	1.00	0.704	0.728	0.136	(0.399, 0.932)	0.4655
p_{11}	0.9	0.21	0.736	0.890	0.292	(0.078, 1.000)	0.9495
p_{22}	0.9	0.21	0.793	0.907	0.252	(0.115, 1.000)	0.8518

Note: Result of 20,000 Gibbs-Sampling iterations, iterations 1,000 through 20,000 used for inference. 95%-band refers to 95% posterior probability bands.

Table 2: Estimation Results for US Merger Activity, 1973:I through 1995:III.

might be that, even if US merger activity between 1973 and 2003 was dominated by a single gigantic tidal wave in the mid 1990s, the assumption of similarly-sized waves downplays the importance of underlying, less gigantic, but nonetheless significant and perhaps more regular wave-like merger activity.

As a straightforward way to investigate this possibility, this section presents estimates for the two-state model using only the data *prior* to the estimated break date which started the tidal wave, i.e. from 1973:I through 1995:III. Note that this corresponds to discarding little more than a quarter of the full series, which should leave us with sufficient data points to identify waves. Table 2 gives the parameter estimates resulting from fitting an AR(1) model. Comparison with estimation results from the full series in Table 1 shows that parameter inference is rather imprecise for the subsample. Figure 5 again shows the inferred probabilities of being in a high activity state for this particular subsample period. By comparison with the clear-cut result presented in Figure 3 for the entire sample period, Figure 5 suggests that a regime switch is rather hard to identify in the merger data up to 1995:III. Although the plot does indeed hint at a somewhat increased probability of a high activity state around the mid 1980s (as well as around the mid 1990s as a warm-up to the ensuing large wave), this hint remains very faint due to the fact that, except for a short period around 1987, the probability of a high activity state stays in a rather tight band around 0.5. Overall, the fact that the inferred probability of a high activity state, $\Pr(S_t = 2|\mathcal{Y}_T)$, stays far clear of either 0 or 1 implies that the

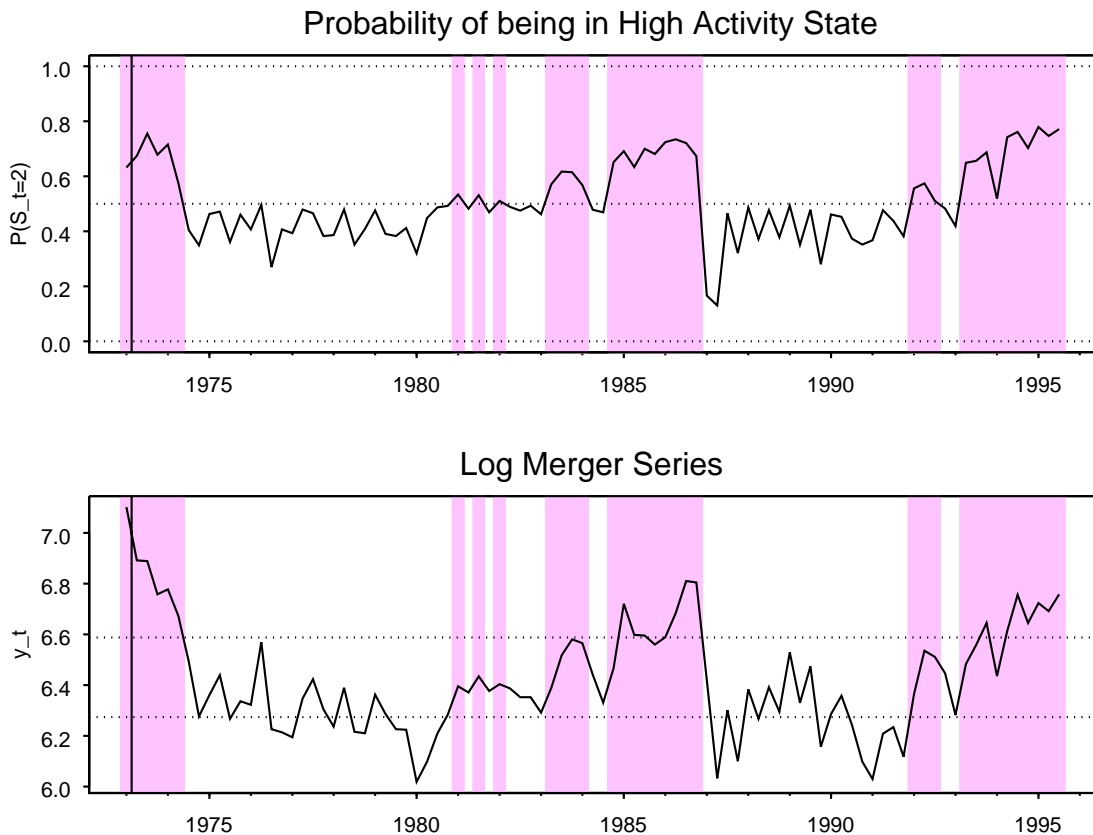


Figure 5: Estimated Probability of Being in State of High Merger Activity, US Log Merger Series 1973:I–1995:III (in Top Panel, Bottom Panel displays Log Merger Series).

data reaches no clear conclusion concerning a likely sequence of states.²⁹

This indecision in state inference is a first clear indication that the US has not witnessed a discrete shift in the mean level of merger activity at all during the 1980s.³⁰ A second striking feature of the probability plot in Figure 5 is its strong qualitative similarity with the underlying time series of log mergers, reproduced in

²⁹An alternative approach would be to extend our two-state model by introducing a third state, thereby explicitly allowing for both ‘medium’ and ‘high’ waves. However, casual inspection of the series strongly suggests that in such a three-state model, all quarters following 1995:III would rather clearly be associated with the ‘high activity state’, leading to inference on ‘low’ vs. ‘medium’ state comparable to the two-state analysis presented in this section.

³⁰Note that in finite samples, given the data \mathcal{Y}_T , $P(S_t = 2|\mathcal{Y}_T)$ will of course generally deviate from 0.5 for any t even if the data were indeed generated by a stationary autoregressive process with no change in regime (i.e., a process with $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$) due to remaining posterior uncertainty about the model parameters. Sample runs of the Gibbs Sampler on simulated data involving no regime change (and parameter values similar to those inferred for US mergers) revealed pictures quite similar to Figure 5.

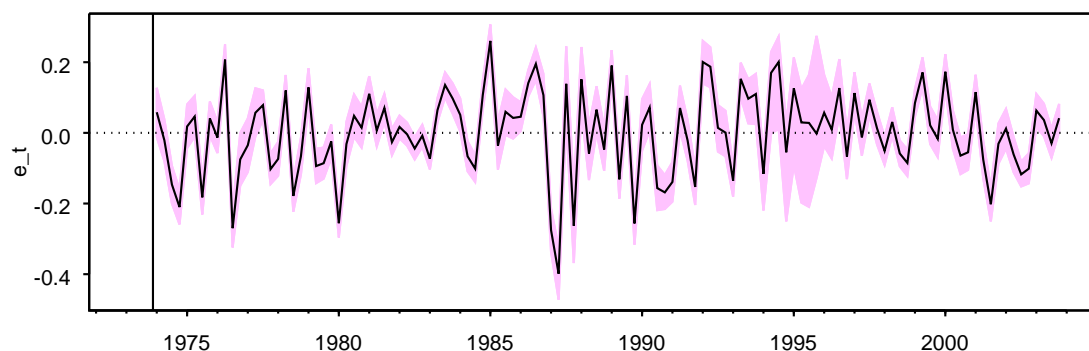


Figure 6: Mean Residuals (with 95% Posterior Bands) for US Log Merger Series 1973:I–2003:IV.

the bottom panel of Figure 5. Indeed, the high-activity estimate in the top panel comes quite close to representing a positive affine transformation of the log merger series in the bottom panel. This is something we should expect to see from fitting a regime switching model to a series generated by a process with no actual switch in regime. Third, Albert and Chib (1993) point out that fitting a Markov switching model to non-switching data results in rather large posterior bands on parameters, particularly concerning the means μ_1 and μ_2 and the transition probabilities p_{11} and p_{22} , which is reflected in our results. Finally and perhaps most importantly, inference runs with simulated data show the posterior distributions of $\mu_2 - \mu_1$ for our US subsample to be quite comparable to the posterior distributions of $\mu_2 - \mu_1$ resulting from simulated series with similar but constant model parameter values (i.e. a series with no state switch).

In sum, therefore, we find little evidence in support of the notion of a discrete 1980s merger wave even in the truncated series. It is important however to be clear about the exact meaning of this result, as it does not contradict of course the 1980s having witnessed somewhat increased merger activity. What our result fleshes out is that this increased activity is rather unlikely to have been associated with a nonlinear shift in regime to the underlying autoregressive process (i.e. a ‘burst in activity’, such as the boom following 1995:III). Rather, it appears to be quite compatible with regular, well-behaved random shocks hitting a stationary linear autoregressive process. This is perhaps best illustrated by Figure 6, which plots the mean residuals (i.e. estimates of ε_t) resulting from the original estimation of the model over the

full data range from 1973:I through 2003:IV.³¹ Overall, these residuals appear well-behaved and compatible with our model assumptions of serial independence and normality. Nonetheless, one might indeed see the first half of the 1980s as having been hit by a sequence of slightly above-average shocks which—amplified by the processes’ strong positive autocorrelation—have given rise to a period of somewhat increased merger activity. We argue, however, that this is compatible with the usual behavior of a stationary autoregressive process rather than signifying a non-linear burst such as a discrete switch in mean over that period. In other words, in terms of our regime switching model of waves, we would like to suggest that the 1980s merger wave symbolized a ‘ripple’ rather than a real wave.

6.4 Parameter Uncertainty Matters: (Mis-)identifying the 1980s Wave

The preceding section has thoroughly discussed our finding that, in contrast to the 1990s merger wave, the increased merger activity in the 1980s constituted no extraordinary burst in activity. This leaves unexplained, however, why the aforementioned studies by Town (1992) and Linn and Zhu (1997), both of which similarly fit a two-state mean-switching Markov model to US merger data, have identified an 1980s merger wave nonetheless.

Candidate reasons for this difference are manifold, as the studies differ in the particular series used, the time span considered, and details in model specification. As this section argues, however, the main difference in our interpretation of the 1980s merger wave is likely to stem from a more subtle, methodological reason: the refined methods of inference offered by the Gibbs sampling approach.

Recall that by means of this approach, our inference on regimes as presented in Figures 3 and 5 is based on $\Pr(S_t = 2|\mathcal{Y}_T)$. This contrasts with Town’s (1992) and Linn and Zhu’s (1997) inference based essentially on $\Pr(S_t = 2|\mathcal{Y}_T; \hat{\beta})$, the likelihood of a high state of merger activity while treating the underlying model parameters as given by the result of a preceding maximum likelihood estimation. Put differently, previous studies have asked the following question: Given that we

³¹Recall that in a Bayesian estimation context, we consider posterior (i.e. ‘updated’) *distributions* of the parameters rather than particular point estimates. Thus, the resulting residuals themselves are random not only due to uncertainty about the unobservable state, but also due to posterior parameter uncertainty. Figure 6 displays both the mean and 95% posterior probability bands for the residual in any period.

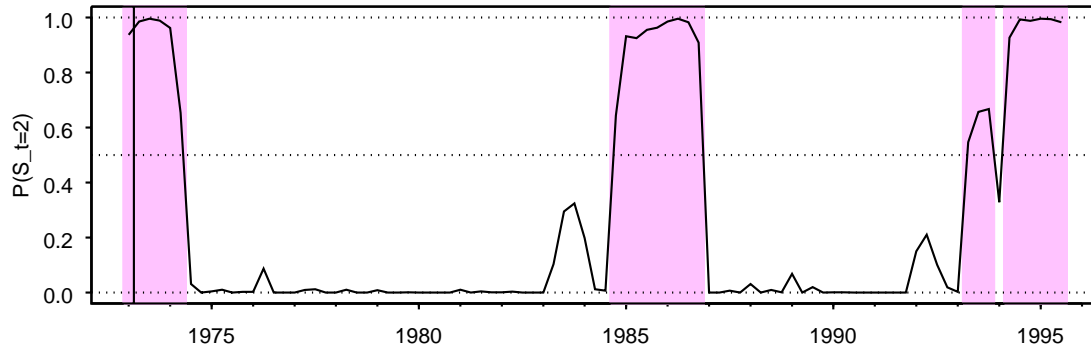


Figure 7: Probability of Being in High State of Merger Activity, US Log Merger Series 1973:I–1995:III, Calculated at Maximum Likelihood Parameter Estimates.

believe the obtained parameter estimates to correspond to the true parameter values, what is the likelihood of a high-activity state in any period? What this question obviously forgets to ask is just how reliable these parameter estimates actually are. Thus, if parameter estimates involve a high degree of uncertainty, answers to this question can severely overstate the evidence in favor of a high state of activity, and this appears to be highly relevant to the discussion of the 1980s merger wave.

To drive this point home, using our data spanning 1973:I through 1995:III, we have replicated the procedure in Town (1992) and Linn and Zhu (1997) by calculating the probability of a high activity state while holding the model parameters fixed at their maximum likelihood estimates reported in Table 2. Figure 7 plots the results. Clearly, by comparison with Figure 5, neglecting parameter uncertainty both leads to considerably more clear-cut inference on regimes and accentuates evidence for a high state of activity in the 1980s. Interestingly, the resulting sharp identification of a merger wave lasting from late 1984 to late 1986 is not very far from findings in Town (1992), who identifies 1986:IV as a period of intense merger activity, and findings in Linn and Zhu (1997), who identify the ‘mid-to-late 1980s’ as a merger wave.

6.5 But Waves Do Exist: The UK Merger Wave Experience

Next, we inspect the UK merger series, shown in the bottom panel of Figure 1, for its wave-like behavior. Analyzing the UK series turns out to be rewarding not only from the point of view of understanding UK merger activity, but also as a

	priors		posteriors				ML
	mean	std	mean	median	std	95%-band	MLE
μ_1	5.00	1000.00	4.808	4.810	0.054	(4.696, 4.913)	4.8188
μ_2	5.00	1000.00	5.838	5.844	0.147	(5.549, 6.123)	5.8625
σ_1	0.15	0.21	0.188	0.188	0.010	(0.170, 0.210)	0.1843
σ_2	0.15	0.21	0.177	0.174	0.026	(0.134, 0.232)	0.1592
ϕ_1	0.00	1.00	0.431	0.430	0.085	(0.266, 0.602)	0.4067
ϕ_2	0.00	1.00	0.189	0.190	0.076	(0.043, 0.338)	0.1743
p_{11}	0.90	0.21	0.981	0.984	0.015	(0.943, 0.999)	0.9811
p_{22}	0.90	0.21	0.915	0.928	0.059	(0.772, 0.988)	0.9119

Note: Result of 20,000 Gibbs-Sampling iterations, iterations 1,000 through 20,000 used for inference. 95%-band refers to 95% posterior probability bands.

Table 3: Estimation Results for UK Merger Activity, 1969:I through 2003:IV.

more general validation of our proposed Markov switching merger model and the wave hypothesis in particular. Indeed, the preceding analysis of the US merger data may be seen as somewhat disappointing as regards the wave hypothesis: While the proposed model itself does seem to provide a very good description of the US data, it does so in a fashion that hardly reflects the *repeated* bursts of activity attributed to mergers by the literature—namely by identifying only a single switch from low to high activity around the mid 1990s, but no subsequent reversion to low activity, let alone a second or even a third wave. Given the limitation of our analysis to the last 30 years of US merger activity, this finding is of course not an invalidation of the wave hypothesis *per se*. However, the idea of a wave-like process governing US mergers would certainly be reinforced by observing more complete and possibly repeated wave cycles in other series such as the UK’s.

To analyze the UK data, we estimate an AR(2)-version of the model in Section 4. Table 3 reports summary statistics on both the priors used and on the estimated posterior distributions of the parameters. As with the full US series, the UK series permits a significant update on the model parameters’ priors, shown by the marginal posteriors’ standard deviation and posterior bands. Again, the inferred high activity mean μ_2 significantly exceeds the low activity mean μ_1 . Interestingly, the ratio of high to low activity (in log terms) seems comparable across the US and the UK, as the high activity mean μ_2 exceeds the low activity mean μ_1 by an average (and mean)

20% for both series.³² Also, shocks to UK mergers again appear to be less volatile in the high activity state, although this is much less significantly so than for the full US series (as would be expected, fitting the model to the untransformed data reveals much higher volatility in the high than in the low activity state). The estimate of ϕ_1 suggests that autocorrelation in the UK series is significantly positive and again quite comparable to the degree observed in the US data. Finally, however, the unobserved state is more likely to change in any period for the UK series, as shown by the posteriors on p_{11} and p_{22} . Particularly, for the UK, the median expected duration of a low activity state is 15.5 years whereas that of a high activity state is 3.2 years. Concerning the ergodic probability of being in a high state (unconditional on the data), $\Pr(S_t = 2)$, the mean posterior is 21.8%, whereas the median is 18.0%.

Next, Figure 8 shows the estimated probabilities of being in a high activity state for the UK merger series. The probability plot shows strong evidence that the UK has witnessed two distinct merger waves between 1969 and 2003. The first wave seems to have lasted from 1971:I through 1973:IV. Due to the aforementioned inference problems at the beginning of the series as well as because the merger series seems somewhat ‘undecided’ prior to 1971, we cannot precisely date the beginning of the first wave. Indeed, our inference leaves open to some extent whether the UK started off in a high or low activity state in 1969.³³ It appears quite clear, however, that this first wave found its end in 1973:IV, as $\Pr(S_t = 2|\tilde{y}_T)$ drops from 0.773 in 1973:IV to 0.035 in 1974:I. The second wave in turn is reliably estimated to have started in 1986:III and ended in 1989:IV (the probability of a high state jumps from 0.372 in 1986:II to 0.987 in 1986:III and dips from 0.957 in 1989:IV to 0.050 in 1990:I).

7 Conclusion

The goal of this paper has been to revisit quantitative evidence on the merger wave hypothesis. Using a model of Markovian parameter switching, recent merger data

³²For an impression of what this means in absolute terms, we note that the median values of $\exp(\mu_1)$ and $\exp(\mu_2)$ are 123 and 345, respectively. Thus, mean activity in the high-activity state exceeds that in the low-activity state by around 180% (the corresponding figure for the US is 260%).

³³In fact, by investigating annual data, Hughes (1993, p. 17) argues that merger activity culminates in twin peaks of activity in the late 1960s and early 1970s.

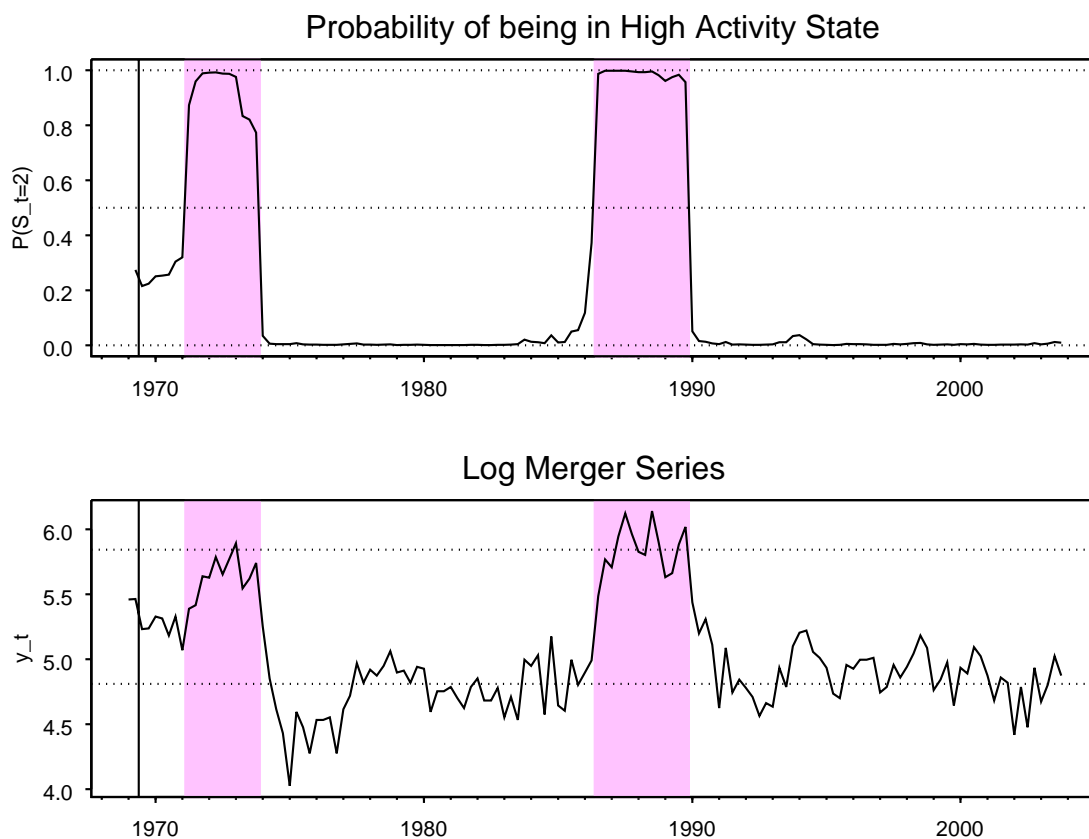


Figure 8: Estimated Probability of Being in State of High Merger Activity, UK Log Merger Series 1969:I–2003:IV (in Top Panel, Bottom Panel displays Log Merger Series).

and refined methods of inference, we have sought to identify and date waves in merger activity. A key finding has been that, concerning merger activity in the US and the UK over the past 30 years, the interpretation of merger activity as a mean and variance switching autoregressive process provides a promising quantitative operationalization of the wave hypothesis.

Moreover, fitting such a model to the data has produced the following merger wave chronology: First, since the beginning of our US series in 1973, the US appear to have witnessed only the beginning of a single large merger wave, this wave having been kicked off between the third and fourth quarter of 1995. Particularly, as a second major result and in contrast to other recent empirical work, we find very little evidence for the much discussed 1980s merger wave—at least in the sense of a discrete shift in average merger activity. We have argued that there is a methodological reason for this discrepancy in findings, as less refined inference methods

which neglect parameter uncertainty are likely to have played a significant role in the econometric misidentification of 1980s merger activity as a wave. Third, fitting our model to UK merger activity between 1969 and 2003 clearly identifies two UK merger waves, one in the early 70s and a second in the late 1980s.

We hope that these findings will serve as a sound basis for a further discussion and investigation of possible underlying causes for merger waves. Particularly, the rather precise identification and precise timing of distinct states of merger activity based on our Markov-switching model openly calls for an economic interpretation and explanation of these states. Investigating one such hypothesis, our brief digression in Section 6.2 concerning industry-level data for the US has argued that, whatever the trigger for the US 1990s wave, industries seem to have been rather uniformly affected by it. Resende (1999), using an industry-level Markov switching model, reaches a similar conclusion concerning UK merger activity. Interestingly, while this leads us to conclude that nationally, industries in each the US and the UK seem to have been similarly affected by the observed waves, comparison of our results in Sections 6.1 and 6.5 quickly reveals that there is no sign of a similar ‘coordination of waves’ across countries over the last 30 years.

References

- Albert, James H., and Siddhartha Chib (1993) ‘Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts.’ *Journal of Business and Economic Statistics* 11(1), 1–15
- Andrade, Gregor, Mark Mitchell, and Erik Stafford (2001) ‘New evidence and perspectives on mergers.’ *Journal of Economic Perspectives* 15(2), 103–120
- Andrews, Donald W. K. (1993) ‘Tests for parameter instability and structural change with unknown change point.’ *Econometrica* 61(4), 821–856
- Avenel, Eric (2008) ‘Strategic vertical integration without foreclosure.’ *Journal of Industrial Economics* 56(2), 247–262
- Barkoulas, John T., Christopher F. Baum, and Atreya Chakraborty (2001) ‘Waves and persistence in merger and acquisition activity.’ *Economics Letters* 70(2), 237–243
- Chow, Gregory C. (1960) ‘Tests of equality between sets of coefficients in two linear regressions.’ *Econometrica* 28(3), 821–856
- (1984) ‘Random and changing coefficient models.’ In *Handbook of Econometrics*, ed. Zvi Griliches and Michael D. Intrilligator, vol. 2 (Amsterdam: North-Holland) pp. 1213–1245
- Chowdhury, Abdur R. (1993) ‘Univariate time-series behaviour of merger activity and its various components in the United States.’ *Applied Financial Economics* 3(1), 61–66
- Clark, John J., Alok K. Chakrabarti, and Thomas Chiang (1988) ‘Trend and stochastic movements in US merger activity.’ *Quarterly Review of Economics and Business* 28(2), 6–19
- Faria, André (2008) ‘Mergers and the market for organizational capital.’ *Journal of Economic Theory* 138(1), 71–100
- Faulí-Oller, Ramon (2000) ‘Takeover waves.’ *Journal of Economics and Management Strategy* 9(2), 189–210
- Gelfand, Alan E., and Adrian F. M. Smith (1990) ‘Sampling-based approaches to calculating marginal densities.’ *The American Statistician* 85(410), 398–409
- Geman, Stuart, and Donald Geman (1984) ‘Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.’ *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6), 721–741

- Golbe, Devra L., and Lawrence J. White (1988) ‘A time series analysis of mergers and acquisitions in the U.S. economy.’ In *Corporate Takeovers: Causes and Consequences*, ed. Alan J. Auerbach (Chicago: University of Chicago Press) pp. 265–305
- (1993) ‘Catch a wave: The time series behavior of mergers.’ *Review of Economic Studies* 75(3), 493–499
- Gärtner, Dennis L. (2007) ‘Why Bayes rules: A note on Bayesian vs. classical inference in regime switching models.’ University of Zürich, SOI Working paper No. 0719
- Gugler, Klaus, Dennis C. Mueller, and B. Burcin Yurtoglu (2005) ‘The determinants of merger waves.’ Working paper, University of Vienna
- Hamilton, James D. (1989) ‘A new approach to the economic analysis of nonstationary time series and the business cycle.’ *Econometrica* 57(2), 357–384
- (1993) ‘Estimation, inference and forecasting of time series subject to changes in regime.’ In *Handbook of Statistics*, ed. G. S. Maddala, C. R. Rao, and H. D. Vinod, vol. 11 (Amsterdam: Elsevier) chapter 9, pp. 231–260
- Harford, Jarrad (2005) ‘What drives merger waves?’ *Journal of Financial Economics* 77(3), 529–560
- Holmstrom, Bengt, and Steven N. Kaplan (2001) ‘Corporate governance and merger activity in the United States: Making sense of the 1980s and 1990s.’ *Journal of Economic Perspectives* 15(2), 121–144
- Hughes, Alan (1993) ‘Mergers and economic performance in the UK: A survey of empirical evidence 1950–1990.’ In *European Mergers and Merger Policy*, ed. Matthew Bishop and John Kay (Oxford: Oxford University Press) pp. 9–95
- Jovanovic, Boyan, and Peter L. Rousseau (2002) ‘The Q-theory of mergers.’ *American Economic Review, Papers and Proceedings* 92(2), 198–204
- Jovanovic, Boyan, and Peter L. Rousseau (2008) ‘Mergers as reallocation.’ *Review of Economic Studies* 90(4), forthcoming
- Kim, Chang-Jin, and Charles R. Nelson (1999) *State-Space Models with Regime Switching* (Cambridge, Mass.: MIT Press)
- Linn, Scott C., and Zhen Zhu (1997) ‘Aggregate merger activity: New evidence on the wave hypothesis.’ *Southern Economic Journal* 64(1), 130–146
- Mitchell, Mark L., and J. Harold Mulherin (1996) ‘The impact of industry shocks on takeover and restructuring activity.’ *Journal of Financial Economics* 41(2), 193–229

- Mueller, Dennis C. (1997) ‘Merger policy in the United States: A reconsideration.’ *Review of Industrial Organization* 12(5–6), 655–685
- Mulherin, J. Harold, and Audra L. Boone (2000) ‘Comparing acquisitions and divestitures.’ *Journal of Corporate Finance* 6(2), 117–139
- Nelson, Charles R., Jeremy Piger, and Eric Zivot (2001) ‘Markov regime switching and unit root tests.’ *Journal of Business and Economic Statistics* 19(4), 404–415
- Nelson, Ralph L. (1959) *Merger Movements in the American Industry, 1895–1956* (Princeton: Princeton University Press)
- Nilsson, Tore, and Lars Sjørgard (1998) ‘Sequential horizontal mergers.’ *European Economic Review* 42(9), 1683–1702
- Owen, Sian (2004) ‘A Markov switching model for UK acquisition levels.’ Working Paper, University of New South Wales, Sydney
- Qiu, Larry D. and Wen Zhou (2007) ‘Merger waves: A model of endogenous mergers.’ *RAND Journal of Economics* 38(1), 214–226
- Ravenscraft, David J. (1987) ‘The 1980’s merger wave: An industrial organization perspective.’ In *The Merger Boom*, ed. L. E. Brown and E. S. Ronsengren Federal Reserve Bank of Boston Conference Series, No. 31, pp. 17–37
- Resende, Marcelo (1999) ‘Wave behaviour of mergers and acquisitions in the UK: A sectoral study.’ *Oxford Bulletin of Economics and Statistics* 61(1), 85–94
- Rhodes-Kropf, Matthew, and S. Viswanathan (2004) ‘Market valuation and merger waves.’ *Journal of Finance* 59(6), 2685–2718
- Scherer, Frederic M., and David Ross (1990) *Industrial Market Structure and Economic Performance* (Boston: Houghton Mifflin Company)
- Shleifer, Andrei, and Robert W. Vishny (1990) ‘The takeover wave of the 1980s.’ *Science* 249(4971), 745–749
- Shleifer, Andrei, and Robert W. Vishny (2003) ‘Stock market driven acquisitions.’ *Journal of Financial Economics* 70(3), 295–311
- Shughart, II, William F., and Robert D. Tollison (1984) ‘The random character of merger activity.’ *RAND Journal of Economics* 15(4), 500–509
- Town, Robert J. (1992) ‘Merger waves and the structure of merger and acquisition activities.’ *Journal of Applied Econometrics* 7, Supplement: Special Issue on Nonlinear Dynamics and Econometrics, S83–S100

Toxvaerd, Flavio (2008) 'Strategic merger waves: A theory of musical chairs.' *Journal of Economic Theory* 140(1), 1–26