Performance persistence in the presence of higher-order resources

Phebo D. Wibbens

Research Summary: This paper presents a formal model of how higher-order resources affect profit persistence. Higher-order resources provide an abstract representation of dynamic capabilities, and are defined as resources that do not affect profit directly, but can affect other resources that in turn affect profit over time. The model shows that higher-order resources lead to persistence not only in the level of profit, but also in its growth. Estimation of the model using empirical profit data of more than 4,000 U.S. firms over 30 years implies an average duration of competitive advantage of about 18 years, which is almost four times as long as implied by traditionally used autoregressive models that exclude the effect of higher-order resources.

Managerial Summary: If firms want to make more profit than their competitors for prolonged periods of time, they must have access to resources that competitors cannot effectively obtain, such as brands, patents, captive customers, or specialized plants. This paper shows that not only these “operating resources” drive long-term profit differences across firms, but also “higher-order resources,” such as strategic planning, M&A capabilities, and superior forecasts. Such higher-order resources do not affect profits directly, but allow firms to obtain superior operating resources over time. A mathematical model incorporating higher-order resources suggests an average duration of competitive advantage of about 18 years, almost four times as long as implied by traditionally used models.

KEYWORDS
formal foundations of strategy, profit persistence, resource based view, dynamic capabilities, Bayesian analysis
1 | INTRODUCTION

The persistence of profitability differences, even among direct competitors, is a well-documented empirical regularity in the strategy literature (e.g., Jacobsen, 1988; McGahan & Porter, 1999; Mueller, 1977). The resource-based view (RBV) provides an explanation for why these profitability differences are not rapidly competed away, invoking firms’ access to heterogeneous and immobile resources with limits to competition. (e.g., Barney, 1991; Peteraf, 1993; Wernerfelt, 1984). The RBV can be augmented with the notion of higher-order resources to explain how certain firms are able to systematically develop and maintain favorable resource positions in their product markets without having to pay their full economic price on strategic factor markets. Unlike “ordinary” or “operating” resources, higher-order resources do not yield more profit directly, but instead allow a firm to systematically obtain superior resources that, over time, allow it to grow profit. This concept of higher-order resources (or higher-order capabilities) provides a way to characterize dynamic capabilities within the RBV (Helfat et al., 2007; Winter, 2003).

However, despite the extensive literature on the RBV and dynamic capabilities, little is known about how higher-order resources theoretically and empirically affect patterns of profit persistence, a particularly puzzling gap given that the relation between resources and profitability has been central to the strategy literature for decades. In this paper, I address this gap with a formal stochastic model of the dynamics between higher-order resources, operating resources, and profits. I find that operating resources, which affect profits directly, lead to persistence in the level of profit differences; the addition of higher-order resources, which affect other resources but not directly profits, leads to persistence in the growth of profit differences.

The model also provides an empirical test of the presence of higher-order resources and their effect on the duration of competitive advantage. A model incorporating the dynamics between higher-order resources, operating resources and profits leads to a so-called second-order autoregressive moving average (ARMA(2,2)) model. Estimating this model on observed profit data implies an average duration of competitive advantage of approximately 18 years. By contrast, commonly used first-order autoregressive (AR(1)) models imply a duration of just 5 years, almost four times as short. The difference stems from two effects. First, the AR(1) model does not take into account higher-order resources, which are accounted for by the second-order autoregressive (AR(2)) term. Second, the AR(1) model does not take into account the fact that annual profit is a noisy measure of underlying resources, which is accounted for by the moving average (MA) terms in the model used here. The results suggest that not accounting for these two effects has likely led to a significant underestimation of the duration of competitive advantage in earlier papers.

This paper makes several contributions to the literature. First, the formal framework captures the precise relations between resources and performance, without equating the two, thus addressing an often-cited tautology critique of the RBV (Priem & Butler, 2001). Resources must persistently affect expected profit over time—providing two critical distinctions between resources and performance. Moreover, resources can be divided into those affecting expected profit directly (operating resources) and those affecting expected profit indirectly, through other resources over time (higher-order resources). The mathematical definitions and derivations in this paper elucidate how the dynamics of operating resources and higher-order resources—individually as well as in conjunction—affect profit persistence patterns in terms of stochastic time series models, as summarized in Table 2.

Second, this study offers a large-scale empirical test of whether and to what extent higher-order resources affect profit persistence patterns. This addresses repeated calls “... to make greater use of empirical methodologies beyond qualitative case analyses and analysis of survey data, such as ...
econometric analysis of ‘big’ archival data, to further broaden the toolkit used in dynamic capabilities research.” (Schilke, Hu, & Helfat, 2018, p. 392). I find that higher-order resources significantly affect profit persistence patterns and the resulting duration of competitive advantage across a wide range of industries. A notable exception is that the effect of higher-order resources on profit persistence patterns is much smaller, or even absent in some industries, in the decade 1995–2004, providing an explanation for the findings of “hypercompetition” literature in the 2000s (e.g., D’Aveni, Dagnino, & Smith, 2010), as well as more recent findings of a reversion of this trend (e.g., Bennett & Gartenberg, 2016).

Third and finally, this paper offers new findings and methods for the persistence of profitability literature (e.g., Bottazzi, Secchi, & Tamagni, 2008; Bou & Satorra, 2007; Geroski & Jacquemin, 1988; Goddard & Wilson, 1999; Jacobsen, 1988; Madsen & Walker, 2017; McGahan & Porter, 1999; Mueller, 1977, 1986; Waring, 1996). For instance, the formal model of resource dynamics implies use of an ARMA model to estimate the time-series behavior of profit persistence instead of the earlier purely autoregressive (AR) models, which may have led to an attenuation bias in prior studies. Indeed, I find significantly higher values for the first-order persistence of profit, around 0.95, compared to earlier values in the range 0.7–0.9. In some industries there are time periods in which the first-order persistence is even above one, consistent with recent findings of nonstationary behavior in profit data (Canarella, Miller, & Nourayi, 2013). Unlike traditionally used maximum likelihood estimation (MLE), the Bayesian inference used here allows estimation of such nonstationary behavior without problems, providing direct estimates of first-order persistence over time. I find that first-order profit persistence above one (i.e., diverging profit differences) never lasts for more than a decade, suggesting the existence of “non-stationary episodes” rather than long-term nonstationary profit behavior.

2 | THEORY

The primary focus of this paper is how the dynamics of operating and higher-order resources affect the persistence of profitability differences among firms. In this section, I define the notions of operating resources and higher-order resources, relate them to previous literature, and mathematically derive how they affect persistence of profits over time as empirically observed.

2.1 | Operating resources

Winter (2003) defines ordinary or “zero-level” capabilities as those “that permit a firm to ‘make a living’ in the short term” (p. 991). The idea of distinguishing capabilities that contribute to short-term profits can logically be extended to other types of resources, which can be defined as “something that the organization can draw upon to accomplish its aims” (Helfat et al., 2007, p. 4). Examples of such resources that directly affect profits include brands, patents, captive customers, or specialized plants. I refer to them as operating resources—the word “operating” is to be interpreted broadly, as in the term “operating routines” (King & Tucci, 2002; Zollo & Winter, 2002)—pertaining not merely to production operations but to any other element of operating in a certain product market. Presumably, such operating resources are those referred to by the original RBV scholars to explain why certain firms are able to make above-normal profits for prolonged periods of time (Barney, 1986; Dierickx & Cool, 1989; Rumelt, 1984; Wernerfelt, 1984). Put succinctly:

Definition 1: Operating resources are resources that directly and persistently affect a firm’s expected profit.
Two words are critical in this definition: “persistently” and “expected.” The latter implies that the effect of a firm’s operating resources $x_t$ in period $t$ on its profit $y_t$ can be described by $x_t = E(y_t)$. Conversely, it implies that firm profit can be written in terms of the effect of its operating resources $x_t$, plus a random noise term $u_t$ that is uncorrelated with the firm’s resource base$^1$:

$$y_t = x_t + u_t$$ (1)

“Persistent” implies that operating resources in subsequent periods must be correlated: $\text{cor}(x_t, x_{t-1}) > 0$. To first-order approximation$^2$ this is equivalent to $x_t = \mu + \lambda x_{t-1} + v_t$, for some parameters $\mu, \lambda > 0$ and some random noise term $v_t$. $^3$ This equation can be simplified by redefining $x_t$ as the difference compared to the competitive average$^4$:

$$x_t = \lambda x_{t-1} + v_t$$ (2)

Since I am primarily interested in the dynamics vis-à-vis the competitive average, rather than the (constant) level of the average itself, from now on I will work with demeaned values: the resource base $x_t$ can be understood as the resource advantage (or disadvantage) vis-à-vis competitors, and the profit $y_t$ can be understood as the profit in excess of the competitive average.

Equation (2) shows the importance of the persistence requirement in the definition. If resources were not persistent, this would mean $x_t = v_t$, and thus $y_t = v_t + u_t$: the profit $y_t$ would be the sum of two uncorrelated noise terms, which is equivalent to a random noise process $y_t = u_t$. Intuitively this makes sense: a short spike in profit due to some very short-lived “resource” is indistinguishable from a spike in profit due to some other random event. Therefore, operating resources are defined as resources that persistently affect (expected) profit, that is, for a prolonged period of time, captured by the requirement $\lambda > 0$. The value of $\lambda$ captures to what extent operating resources are persistent.

As will be shown later, Equations (1) and (2) are equivalent to a so called first-order AR and first-order moving average (ARMA(1,1)) model for profits $y_t$. Previous literature has usually just considered a first-order AR (AR(1)) model. However, it turns out that the omission of the MA term is not inconsequential, and has likely led to the under-estimation of profit persistence in earlier studies.

### 2.2 Higher-order resources

Analogous to extending the concept of “zero-level” capabilities to other resources that directly affect profit, dynamic or “higher-order” capabilities (Winter, 2003) can be extended to other resources that govern the “rate of change” of operating resources (Collis, 1994; Winter, 2003):

**Definition 2:** Higher-order resources are resources that do not directly affect current expected profit, but that do persistently affect the expected rate of change of other resources.

Similar to the definition of operating resources, the notions of persistence and expected values are critical. As Winter (2003, p. 992) notes, the requirement of such patterns beyond random noise is needed to determine whether higher-order resources exist in some nontrivial sense: if higher-order

---

$^1$More formally: $\text{cor}(u_t, x_t) = \text{cor}(u_t, u_{t-1}) = E(u_t) = 0$.

$^2$The persistence of profit literature usually assumes an autoregressive model for profits. All these models are linear time series models: excess profits $y_t$ in a given year are described in terms of a linear combination of its lags ($y_{t-1}, y_{t-2}, \ldots$) as well as an innovation $\epsilon_t$ and potentially its lags ($\epsilon_{t-1}, \epsilon_{t-2}, \ldots$). In order to keep the model mathematically tractable as well as to make it comparable with the earlier literature, I want to stay within this framework of linear time series models, and thus require linear relations between profit and resources.

$^3$Similarly: $\text{cor}(v_t, u_t) = \text{cor}(v_t, v_{t-1}) = \text{cor}(v_t, v_s) = E(v_t) = 0$.

$^4$Formally, this amounts to shifting $x_t$ and $y_t$ by a constant, redefining $x_t \rightarrow x_t + \mu$ and $y_t \rightarrow y_t + \mu$. In the case $\lambda = 1$ one can add or subtract an arbitrary constant without affecting the dynamics.
resources were simply defined as the rate of change of operating resources, they would be trivial, because evidently firms’ resource positions change all the time. The interesting question is whether such change is patterned, and hence the rate of change of operating resources persists over time. The primary purpose of this paper is therefore to derive the effect of such persistent higher-order resources on the time series behavior of profit, as well as whether and how broadly this effect can be empirically observed.

To mathematically formalize the effect of higher-order resources on profit persistence, I only consider second-order resources, that is, resources that directly affect operating resources.\(^5\) Conceptually, the argument can be easily extended to third-order resources (resources that affect the rate of change of second-order resources), and beyond.

By definition, and analogous to operating resources, higher-order resources \(z_t\) can be described by their effect on the rate of change of the operating resources, beyond the baseline from Equation (2):

\[
x_t = \lambda^{(1)} x_{t-1} + z_t + v_t \tag{3}
\]

In analogy to the evolution of operating resources, the evolution of higher-order resources \(z_t\) itself is described by:

\[
z_t = \lambda^{(2)} z_{t-1} + w_t \tag{4}
\]

Summarizing, three equations describe the model:

\[
\begin{align*}
\text{Annual profit:} & \quad y_t = x_t + u_t & & (1) \\
\text{Operating resources:} & \quad x_t = \lambda^{(1)} x_{t-1} + z_t + v_t & & (3) \\
\text{Higher-order resources:} & \quad z_t = \lambda^{(2)} z_{t-1} + w_t & & (4)
\end{align*}
\]

In these equations, \(u_t, v_t,\) and \(w_t\) are independently distributed random variables, while \(\lambda^{(1)}\) and \(\lambda^{(2)}\) capture the persistence of, respectively, operating and higher-order resources. In technical terms, this constitutes a linear state-space model: \(x_t\) and \(z_t\) are the state variables, \(y_t\) is the output variable, Equations (3) and (4) are the transition equations, and Equation (1) is the emission equation.

In the next subsection I use several numerical examples to show how the \(\lambda\)-parameters affect profit persistence, and then more formally derive the implied time-series patterns.

### 2.3 Numerical examples

To intuitively understand the model in Equations (1), (3), and (4), Table 1 shows several numerical examples for various values of \(\lambda^{(1)}\) and \(\lambda^{(2)}\). For all examples, the initial stock of both operating and higher-order resources \(x_0 = z_0 = 100\). I suppress the random shocks \(u_t = v_t = w_t = 0\) in order to focus exclusively on the effect of resource persistence on profits—in the more formal treatment in the next subsection I include the random shocks to precisely derive their impact on the time-series behavior of profit.

Panel (a) shows the evolution of profit \(y\), operating resources \(x\) and higher-order resources \(z\) in the case where operating resources fully persist \((\lambda^{(1)} = 1)\), while higher-order resources decay

---

\(^5\)The terminology for numbering of the higher orders in the literature is inconsistent; for instance, what I call “second-order resources” corresponds to “first-order dynamic capabilities” in Winter (2003) and to “second-order competences” in Danneels (2016).
immediately ($\lambda^{(2)} = 0$); thus, only operating resources affect profit evolution in this example. In this simple case, operating resources, and thus profit levels persist, over time.

Panel (b) shows what happens if the roles of operating and higher-order resources are interchanged: higher-order resources fully persist ($\lambda^{(2)} = 1$), while operating resources decay immediately ($\lambda^{(1)} = 0$). Interestingly, this leads to the exact same pattern with constant profit. Though maybe surprising at first, this can be understood intuitively: the profit pattern derived directly from a stable operating resource is equivalent to the profit pattern derived via a short-lived operating resource from a stable higher-order resource.

Panel (c) shows the pattern in the case where both operating and higher-order resources fully persist ($\lambda^{(1)} = \lambda^{(2)} = 1$). This leads to a markedly different pattern of profit evolution: now not only does the profit level persist, but also profit growth persists. Thus, the interaction of operating and higher-order resources leads to a unique pattern that is distinct from that which would be possible if only operating or higher-order resources were present.

Panel (d) shows the evolution when resource advantages erode over time. Specifically, $\lambda^{(1)} = 0.5$, meaning that operating resources on average decline by 50% per year, and $\lambda^{(2)} = 0.9$, meaning that higher-order resources on average decline by 10% per year. This leads to a decline in $z_t$ from a value of 100 in year 0 to 59 in year 5. The evolution of the operating resource advantage $x_t$ (which equals

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Numerical examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a. Operating resources ($\lambda^{(1)} = 1, \lambda^{(2)} = 0$)</strong></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>100</td>
</tr>
<tr>
<td>$x$</td>
<td>100</td>
</tr>
<tr>
<td>$z$</td>
<td>100</td>
</tr>
<tr>
<td><strong>b. Higher-order resources ($\lambda^{(1)} = 0, \lambda^{(2)} = 1$)</strong></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>100</td>
</tr>
<tr>
<td>$x$</td>
<td>100</td>
</tr>
<tr>
<td>$z$</td>
<td>100</td>
</tr>
<tr>
<td><strong>c. Both resources ($\lambda^{(1)} = \lambda^{(2)} = 1$)</strong></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>100</td>
</tr>
<tr>
<td>$x$</td>
<td>100</td>
</tr>
<tr>
<td>$z$</td>
<td>100</td>
</tr>
<tr>
<td><strong>d. Eroding resources ($\lambda^{(1)} = 0.5, \lambda^{(2)} = 0.9$)</strong></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>100</td>
</tr>
<tr>
<td>$x$</td>
<td>100</td>
</tr>
<tr>
<td>$z$</td>
<td>100</td>
</tr>
<tr>
<td><strong>e. Eroding resources ($\lambda^{(1)} = 0.9, \lambda^{(2)} = 0.5$)</strong></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>100</td>
</tr>
<tr>
<td>$x$</td>
<td>100</td>
</tr>
<tr>
<td>$z$</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. Numerical examples for the model in Equations (1), (3), and (4), showing the evolution of profit $y_t$, operating resources $x_t$, and higher-order resources $z_t$ over time for different values of $\lambda^{(1)}$ and $\lambda^{(2)}$. For all examples $x_0 = z_0 = 100$ and $u_t = v_t = w_t = 0$. 

...
excess profits $y_t$ (in the absence of random shocks) is shaped by two opposing forces, the positive higher-order resource advantage $z_t$ drives growth in operating resources, while the direct effect due to $\lambda^{(1)} < 1$ drives a decline. In the first 2 years, the first force is stronger, while thereafter the second become stronger as the advantage of higher-order resources diminishes.

Panel (e) shows an example similar to panel (d), but with the values of $\lambda^{(1)}$ and $\lambda^{(2)}$ interchanged. Interestingly, this does not affect the evolution of the operating resource advantage $x_t$ and excess profit $y_t$ (though the evolution of higher-order resources $z_t$ is different). This is another instance of the equivalence observed between panels (a) and (b). In general, for the time-series behavior of profits it does not matter whether a firm possesses a rapidly deteriorating operating resource advantage and a slowly deteriorating higher-order resource advantage, or vice versa.

These examples suggest, first, that profit behavior in the presence of both operating and higher-order resources is markedly different than in the presence of either one alone, and second that the roles and persistence rate of operating and higher-order resources can be interchanged without affecting profit persistence patterns. Below I will derive these statements more formally.

### 2.4 Transformation into ARMA time series

The persistence of profitability literature often uses AR models to describe the time-series behavior of profits (e.g., Bottazzi et al., 2008; Bou & Satorra, 2007; Geroski & Jacquemin, 1988; Goddard & Wilson, 1999; Jacobsen, 1988; Madsen & Walker, 2017; McGahan & Porter, 1999; Mueller, 1977, 1986; Waring, 1996). These time-series models are part of a larger class of ARMA models. ARMA models have been extensively studied and used in various scientific disciplines because of their parsimony, wide-ranging applications, and well-known properties (e.g., Cowpertwait & Metcalfe, 2009; Hamilton, 1994).

AR processes describe $y_t$ in terms of the sum of lagged variables up to a certain lag $p$, plus a random shock: $y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t$. Because the random variable $y_t$ at time $t$ feeds into the equation for the next year, AR processes have a “long-term memory”: even for an AR process with just one time lag (AR(1)), a random shock $\epsilon_t$ in a certain year can affect many future years through various time lags. These properties, along with the fact that AR(1) models can easily be estimated with ordinary least squares (OLS) (by regressing a variable on its first lag), make them popular for time series analyses, for instance in the earlier mentioned persistence of profitability literature.

MA processes describe $y_t$ in terms of the sum of lagged error terms up to a certain lag $q$: $y_t = \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$. In contrast to AR models, MA models have a “short-term memory”: any random shock can only affect the outcome variable up to the maximum lag of $q$ time periods later, after which it literally drops out of the equation. Thus MA processes provide a natural model for noise with a short-lived impact.

Because the linear model in Equations (1), (3), and (4) consists of a combination of AR processes (the decay of resource [dis-]advantages $x_t$ and $z_t$) and short-lived noise terms (e.g., the annual variation $u_t$), it might be expected that the implied time-series behavior of $y_t$ consists of a combination of AR and MA terms. This is indeed the case:

**Proposition 1:** Assume the stochastic model as defined in Equations (1), (3), and (4). Then the time series $y_t$ is second-order equivalent to an ARMA(2,2) process with zero mean. In other words, there exist identically and independently distributed $\epsilon_t$, called the innovations, and parameters $\theta_k$, $\phi_k$ such that the time series $y_t$ defined by

6General autoregressive and moving average processes can also include a constant term (e.g., $y_t = c + \phi y_{t-1} + \epsilon_t$ for an AR(1) process). Because in this paper I work with demeaned variables, I omit the constant.
\[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \]  

(5)

has the same expected values and variance–covariance structure as the time series \( y_t \) defined by Equations (1), (3), and (4). Moreover, the values of \( \lambda^{(1)} \) and \( \lambda^{(2)} \) determine the AR parameters \( \phi_1 \) and \( \phi_2 \) as follows:

\[ \phi_1 = \lambda^{(1)} + \lambda^{(2)} \]  

(6)

\[ \phi_2 = -\lambda^{(1)} \lambda^{(2)} \]  

(7)

Proof: See Appendix A of Supporting Information Appendix S1.

Thus, Equations (1), (3), and (4) describe an ARMA(2,2) model for \( y_t \). Moreover, from Lemma 1 in Appendix A of Supporting Information Appendix S1 it follows that a model with only operating resources and no higher-order resources (i.e., \( \lambda^{(2)} = 0 \)) is equivalent to an ARMA(1,1) model. A great advantage of the ARMA formulation over the original one is that the ARMA estimation properties are well known: the model is identified as long as there are no jointly redundant AR and MA components present in the model (e.g., Hamilton, 1994).

Intuitively, the relation between the earlier model and the ARMA formulation can be understood as follows. In a model with no higher-order resource advantages \( (z_t = w_t = \lambda^{(2)} = 0) \) and no annual noise term \( (u_t = 0) \), the excess profit \( y_t \) follows an AR(1) model, because in that case \( y_t = x_t \) and thus \( y_t = \lambda^{(1)} y_{t-1} + \nu_t \). The addition of the annual noise term \( u_t \) essentially means that the excess profits \( y_t \) become a noisy measure of the operating resource advantage \( x_t \), which leads to the addition of a MA term to the time-series model: ARMA(1,1). The addition of higher-order resource advantages leads to both an additional AR term and an additional error term, which translates into a second MA term, thus defining an ARMA(2,2) model.

Note that \( \lambda^{(1)} \) and \( \lambda^{(2)} \) are fully interchangeable in the sense that exchanging them leads to the exact same AR parameters, consistent with the examples provided in Table 1 panels (d) and (e). Therefore, in the empirical inference one needs to decide which value to assign to which \( \lambda \)-parameter in order to uniquely determine them. For the remainder of the paper I will assign the lower value to \( \lambda^{(2)} \), as this provides a lower bound of the persistence rate of higher-order resource advantages, and thus a more conservative estimate of this value.

Table 2 summarizes the results from this section. If no operating and higher-order resources are present \( (\lambda^{(1)} = \lambda^{(2)} = 0) \), then profit evolution is equivalent to \( y_t = u_t \), a random walk, without any AR components. If only operating resources are present \( (\lambda^{(1)} > 0, \lambda^{(2)} = 0) \), then profit levels persist over time, leading to a first-order AR component. Because of the interchangeability of \( \lambda^{(1)} \) and \( \lambda^{(2)} \), the case with only higher-order resources present \( (\lambda^{(2)} > 0, \lambda^{(1)} = 0) \) is exactly equivalent to the case with only operating resources. Finally, presence of both operating and higher-order resources leads to the addition of higher-order AR and MA terms, described by a model with two AR components;

<table>
<thead>
<tr>
<th>Presence of resources</th>
<th>Persistence of profit</th>
<th>Time series Number of AR components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level</td>
<td>Growth</td>
</tr>
<tr>
<td>Operating</td>
<td>Higher-order</td>
<td>--</td>
</tr>
<tr>
<td>✓</td>
<td>--</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

TABLE 2  Summary of the effects of operating and higher-order resources on profit persistence
intuitively this can be understood as the persistence of not only profit levels but also profit growth over time, as per the example in Table 1 panel (c).

These results provide an empirical test for the presence of higher-order resources in shaping profit persistence. In a world with only operating resources, one would expect excess profits to follow an ARMA(1,1) process, while in a world in which higher-order resources also play a significant role, one would expect higher-order AR components to be present in profit time series. Note that, while the AR terms determined by the $\lambda^{(k)}$ parametrize the long-term profit persistence patterns, the MA parameters $\theta_k$ only parametrize short-term noise. Therefore the MA parameters are typically treated as “nuisance parameters” in this type of model (Harris, 1999, p. 158), analogous to, for instance, the $SD$ of the error term in OLS: such parameters are needed in the model for correct inference, but their values are usually not of interest by themselves.

To provide an intuitive understanding for the impact of the second order AR term on long-term profit patterns, Figure 1 shows two simulations, first of an ARMA(1,1) time series, and second of the same model with the addition of an AR(2) term. Similar to the stylized examples in Table 1, the addition of the AR(2) term leads to markedly different behavior. With the addition of the AR(2) term, profit changes become persistent over time, leading to larger and longer deviations from the competitive average. The remainder of this paper will be devoted to measuring the presence of the

![ARMA time series simulations](image)

**FIGURE 1** ARMA time series simulations. *Note: Simulations of ARMA time series models with $\lambda^{(1)} = 0.9$, $\theta_1 = -0.7$, and a standard normal distribution for the innovations $\epsilon_t$; in the ARMA(2,1) model a second-order component with $\lambda^{(2)} = 0.7$ has been added.*
AR(2) term as well as the magnitude of both AR components in observed profit data, and thus of the importance of operating and higher-order resources in shaping long-term profit patterns.

3 | METHODS

3.1 | Data and measures

To determine whether higher-order AR components are present in actual profit time series, I use the Compustat North-America database—the de facto standard for this type of analysis—over the 30-year period from 1985 to 2014 (inclusive). Consistent with other studies using performance data (e.g., McGahan & Porter, 1997; Ruefli & Wiggins, 2003; Rumelt, 1991), for each firm $i$ in industry $j$ and year $t$, I measure assets $K_{ijt}$ through total balance sheet assets (Compustat field code at in Wharton Research Data Services), and profits $\pi_{ijt}$ through operating income (Compustat field code oiadp; also known as earnings before interest and taxes). I operationalize industry definitions through six-digit GICS codes (as a robustness check, I also use SIC, see Appendix C of Supporting Information Appendix S1; I use GICS codes for my main specification, because they provide more contemporary industry definitions than the decades-old SIC codes, and lead to more balanced distributions of firms across industries). All monetary figures are deflated to 2010 real U.S. dollars using annual consumer price index figures. The excess profits $y_{ijt} = \pi_{ijt} - r_{jt}K_{ijt}$ are calculated for each company-year using the industry average returns for each industry-year $r_{jt}$.

Note that the measure of excess profits differs by a size factor from the commonly used return on assets (RoA) vis-à-vis competitors. Specifically, denoting a firm's RoA with $\bar{\pi}_{ijt} = \frac{\pi_{ijt}}{K_{ijt}}$, my measure of excess profits is $y_{ijt} = (\bar{r}_{ijt} - r_{jt})K_{ijt}$: the firm's RoA above or below the industry average, times the firm's assets. This measure is closely related to the notion of economic profit, replacing the cost of capital (which is on average equal to the average returns of all firms), by an industry-year average return. The reason for using this measure is that firms ought to optimize a dollar amount of excess profits, instead of a ratio such as RoA (Levinthal & Wu, 2010). For related reasons, Hawawini, Subramanian, and Verdin (2003) use economic profit instead of RoA to analyze profit patterns.

In the sample, I exclude all financial services companies because of their different accounting standards, as well as all observations with assets below 100 million real 2010 U.S. dollars, to limit the sample size (and thus computing time) to the economically most relevant companies. Since I exclude any observations where there has been a missing observation for a specific company, I obtain an unbalanced panel with contiguous financial data per company. Finally, I exclude companies with fewer than 10 observations and all industries that have fewer than five companies in some years. This leads to a sample of 79,203 observations by company-year for 4,321 companies in 57 industries. In Appendix C of Supporting Information Appendix S1, I check for the robustness of the results to the specific cut-off points used.

3.2 | Box–Jenkins method

Using the Compustat data, I want to estimate the parameters of the model defined in Equations (1), (3), and (4), or equivalently the parameters of the ARMA transformed model in Equation (5). However, an ARMA model in general cannot be estimated using regression techniques such as OLS (Harris, 1999, p. 149). Instead, the Box–Jenkins method is the canonical approach for time-series analysis (Box & Jenkins, 1970; Hamilton, 1994). This method relies on an analysis of (partial) autocorrelation functions and MLE of the ARMA parameters $\theta_k$ and $\phi_k$ in Equation (5).
The autocorrelation function \( \rho_s = \text{Cov}(y_t, y_{t-s})/(\sigma(y_t)\sigma(y_{t-s})) \) is simply the correlation between time lags \( s \) of the time series \( y_t \). The autocorrelation function can be estimated by just taking the sample autocorrelations. The partial autocorrelation function \( \alpha_s \) is the correlation between the \( y_t \) and \( y_{t-s} \) after projecting out linear combinations time series \( y_t \) in between times \( t \) and \( t-s \). It can be estimated by taking a regression of \( y_t \) on the time series \( y_{t-1}, \ldots, y_{t-s} \); the partial slope of \( y_{t-s} \) provides an estimate of the partial autocorrelation \( \alpha_s \) (Hamilton, 1994, p. 111). These functions have standard implementations in the major statistical packages. I will use the implementation in R, using the functions `acf` and `pacf`, respectively (R Core Team, 2016).

The (partial) autocorrelation functions can be used as a first model-free assessment of the time-series structure. For instance, for a pure MA(\( q \)) process, the autocorrelation function \( \rho_s \) is zero for \( s > q \), while for a pure AR(\( p \)) process, the partial autocorrelation \( \alpha_s \) is zero for \( s > q \). If both \( \rho_s \) and \( \alpha_s \) have several significant terms, this usually indicates a mixed ARMA(\( p, q \)) process.

The initial assessment using (partial) autocorrelation functions can then be used to estimate specific ARMA models using MLE, providing confidence intervals for the ARMA parameters \( \theta_k \) and \( \phi_s \). In R this estimation procedure is implemented in the `arima` function. The significance of the terms and the quality of model fit (assessed through, e.g., the Akaike Information Criterion, AIC) also provide an opportunity for selecting the best ARMA model. Moreover, the innovations \( \epsilon_t \) that remain after fitting the ARMA model can again be tested for (partial) autocorrelations—if these are still significant, this indicates the need for a higher-order ARMA model.

One problem with the Box–Jenkins approach in this setting is that it is designed for a single time series \( y_t \), while in this analysis there is a separate time series \( y_{ijt} \) for each firm \( i \). In order to still employ the Box–Jenkins approach, I analyze all time series together as if they were generated from a single ARMA model. I implement this by dividing each individual time series \( t \mapsto y_{ijt} \) by its sample SD (so each time series is at the same scale) and then inserting a sufficient number of “missing” observations between the different time series of each individual firm. For instance, when making assessments of up to lag five in either the (partial) autocorrelation functions or the MLE, at least five “missing data” points are needed in order to make sure that estimates are not contaminated between time series from different companies. However, this approach still makes the assumption that all ARMA parameters are the same across industries, which is not plausible given earlier studies (e.g., Waring, 1996).

### 3.3 Bayesian inference

Hierarchical modeling using Bayesian estimation can overcome the potentially troublesome assumption that all parameters are equal across industries, while still obtaining full statistical power from all data (Alcácer, Chung, Hawk, & Pacheco-de Almeida, 2018; Gelman et al., 2013, ch. 5). In the Bayesian framework, parameters are modeled as a vector of random variables \( \theta \). The statistical model consists of two parts: the prior distribution \( p(\theta) \) and the likelihood \( p(y|\theta) \) of the observations \( y \) given certain parameters \( \theta \). Bayesian inference then proceeds by assessing the posterior \( p(\theta|y) \propto p(y|\theta)p(\theta) \), which describes the probability distribution of the parameters given the actually observed data. This contrasts with the classical MLE, which merely describes the mode (i.e., maximum) of the likelihood function \( \theta \mapsto p(y|\theta) \), while the standard errors provide an approximation of the width of the peak of the likelihood function. Thus, when taking a flat prior \( p(\theta) \propto 1 \) in Bayesian inference, the posterior mode and second derivative are equal to the MLE. Hence, MLE (and thus regression) can be seen as a specific case of Bayesian inference.

Because parameters in Bayesian inference are modeled stochastically, parameter values for different industries can be assumed to be draws from some common probability distribution—the
hierarchical prior. Usually, such hierarchical models provide more reasonable parameter estimates than nonhierarchical models (Gelman et al., 2013, p. 101). For instance, the persistence parameters $\lambda_j^{(k)}$ for each industry $j$ can be drawn from a normal distribution:

$$\lambda_j^{(k)} \sim \mathcal{N}(\mu_{\lambda,k}, \sigma_{\lambda,k})$$ (8)

The AR parameters $\phi_k$ for each draw can be calculated from the persistence parameters $\lambda^{(k)}$ using Equations (6) and (7).

The $SD$s $\sigma_{\lambda,k}$ in expression (8) provide estimates of the spread in parameters across industries, thus giving an estimate of the extent to which resource dynamics differ across industries. Such measures of variance can be interesting measures for strategic research, and are hard to obtain using classical techniques such as OLS and MLE (Alcácer et al., 2018).

The advantage of Bayesian analysis for hierarchical modeling comes at the disadvantage of a much bigger computational burden to calculate the posterior distribution than calculating the classical OLS, maximum likelihood (ML), or generalized method of moments (GMM) estimates (which essentially are approximations of a full Bayesian posterior). Because the posterior is analytically intractable except for the simplest models, it usually needs to be simulated: hundreds or thousands of randomized instances of the model are calculated, which represent draws from the posterior distribution. Inference then proceeds with calculation of statistics from these draws, such as the mean, the median, the $SD$ and the 95% intervals—this is the posterior inference. The posterior mean or median corresponds to classical point estimates such as the MLE. The posterior $SD$ and 95% interval correspond to the classical standard error and 95% confidence interval, respectively. The central limit theorem guarantees that for large $N$ all distributions are normal, with the same parameters using either Bayesian or classical inference (Gelman et al., 2013, ch. 4).

With the advent of modern information technology (IT), the computational costs of such simulations have greatly diminished. For instance, estimating the hierarchical model used in this paper on close to 100,000 observations takes up to a few days, using a parallelized algorithm on designated hardware. Thus, in this case, the computational burden is manageable, while the benefits of being able to model parameters by industry are significant.

Appendix B of Supporting Information Appendix S1 describes the full specification of the Bayesian model used in this paper. It combines the ARMA transformation of the original model as outlined in Proposition 1 with the hierarchical model as outlined in this section. The prior $p(\theta)$ is designed to be minimally informative: all posterior information comes from the data. Note that because of the large data set used to estimate the model, the specific choice of prior should have minimal influence on the posterior distribution in any case.

4 | RESULTS

4.1 | Descriptive statistics, autocorrelations, and ML estimates

Since I perform a time-series analysis, there is only one variable of interest: the excess profits $y_{ijt}$ for each company $i$ in industry $j$ and year $t$, measured in million 2010 U.S. dollars. By definition, the average of $y_{ijt}$ is zero (in fact, the average for each industry-year equals zero, because each company's profits are measured compared to the industry average in each year). Note that due to the highly skewed size distribution of companies, most companies make a small profit or loss compared to the
average (half of the companies between −46 and +43 million), while a few companies make a large profit or loss (about 5% of companies more extreme than plus or minus 1 billion).

In order to bring the distribution across different companies to the same scale when they are analyzed as a single joint time series, each time series is divided by its standard deviation ($y_{ij}/SD[y_{ij}] \rightarrow y_{ij}$). Moreover, this greatly reduces the influence of any outliers: if a particular time series contains an extreme value, the $SD$ for this company will be large, and the absolute scale of this time series thus strongly diminished. Note that this transformation does not change the time-series pattern (i.e., the $\lambda$ and $\theta$ parameters); it merely rescales each time series by a constant. All analyses are performed on this rescaled variable $y$.

Table 3 lists the autocorrelation and partial autocorrelation function of $y_{ij}$ for various time lags. Given that $N = 79,203$, values more extreme than $\pm 2/\sqrt{79,203} \approx \pm 0.007$ are significantly different from zero at the 95% level (Hamilton, 1994, p. 111). Clearly, all autocorrelations and partial autocorrelations up to lag 5 are significant, suggesting the presence of both AR and MA components in the data.

The ML estimates in Table 4 corroborate this picture. A comparison of both the AIC and the parameter estimates indicates that at least two AR components and one MA component are highly significant. The AIC is lowest for the ARMA(2,2) model—indicating the best model fit for that specification. Moreover, the $\lambda^{(2)}, \theta_1$ and $\theta_2$ estimates are significantly different from zero in each of the models, further suggesting that the ARMA(2,2) model provides the best ML fit for this time series.

---

### Table 3 (Partial) autocorrelation functions

<table>
<thead>
<tr>
<th>Lag (years)</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.798</td>
<td>0.798</td>
</tr>
<tr>
<td>2</td>
<td>0.677</td>
<td>0.112</td>
</tr>
<tr>
<td>3</td>
<td>0.601</td>
<td>0.087</td>
</tr>
<tr>
<td>4</td>
<td>0.547</td>
<td>0.066</td>
</tr>
<tr>
<td>5</td>
<td>0.510</td>
<td>0.064</td>
</tr>
</tbody>
</table>

*Note. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of excess profits $y$."

### Table 4 Model comparison for MLE

<table>
<thead>
<tr>
<th></th>
<th>AR(1)</th>
<th>ARMA(1,1)</th>
<th>ARMA(2,1)</th>
<th>ARMA(2,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda^{(1)}$</td>
<td>0.809</td>
<td>0.867</td>
<td>0.928</td>
<td>0.925</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\lambda^{(2)}$</td>
<td></td>
<td>0.477</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.024)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>−0.172</td>
<td>−0.708</td>
<td>−0.626</td>
<td></td>
</tr>
<tr>
<td>(0.005)</td>
<td>(0.011)</td>
<td>(0.026)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td></td>
<td></td>
<td>−0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>AIC</td>
<td>206,707.8</td>
<td>205,637.7</td>
<td>204,878.2</td>
<td>204,866.0</td>
</tr>
</tbody>
</table>

*Note. Maximum likelihood estimates for various autoregressive moving average (ARMA) models of company excess profits. Standard errors in brackets. The Akaike information criterion (AIC) provides an estimate of model fit: lower values indicate a better fit to the data, accounting for the number of parameters in the model.

---

The (partial) autocorrelation functions are calculated in R for a single time series $y_i$, which consists of all time series $y_{ij}$ for each firm $i$, with a sufficient number of NA values inserted between the firms’ time series, so that the autocorrelations only reflect the within-firm values, and not spurious correlations between time series from different firms.
The $\lambda^{(1)}$ estimate for the AR(1) model is in line with the values commonly reported in the literature, which are usually in the range 0.6–0.9 for models without firm fixed effects (e.g., Bottazzi et al., 2008; Waring, 1996, Figure 3). This value is significantly lower though than the $\lambda^{(1)}$ value for the other models, particularly ARMA(2,1) and ARMA(2,2). This suggests that the reported values in the literature are too low as a result of the misspecification of the model—that is, leaving out the MA and AR(2) components. Indeed, leaving out a MA term should lead to attenuation of the parameter estimates. Essentially, when using profitability in a regression to determine the AR(1) coefficient, this introduces a measurement error due to annual variations that are independent of the “true” underlying resource position. The resulting measurement error in the independent variable is a source of endogeneity and creates an attenuation bias in the parameter estimates (e.g., Wooldridge, 2010). Hence it should be expected that the reported values in the literature are lower than the actual persistence, consistent with the findings in Table 4.

I also analyzed the autocorrelations in the resulting $\epsilon_t$ from the various ML inferences. The error terms from the ARMA(1,1) model still have significant (partial) autocorrelations for first and second-order lags of 0.025 and −0.056 respectively, well outside the 95% confidence interval of ±0.007 expected from a null model, providing further evidence for higher-order components present in the data. The (partial) autocorrelations of the ARMA(2,2) model are much smaller, suggesting that those models account for most or all of the correlation structure in the data.

Finally, the normal probability plots for the error terms exhibit roughly a straight line, suggesting that the MLE does not suffer from outliers or other issues as a result of deviations from the normal distribution.

### 4.2 Bayesian inference

To better account for variations across industries, I also performed a hierarchical Bayesian estimation as described in the Methods section. All results are based on posterior inference of 4 Hamilton Monte Carlo (HMC) chains of 5,000 simulations each. The first 2,500 simulated draws were discarded as warm-up,8 and the remainder was thinned by a factor 20 (to conserve memory and disk space), for a total of 500 posterior draws for each model. Due to correlations between the different draws, the effective sample size is sometimes lower. For each parameter of interest, $\hat{n}_{\text{eff}}$ indicates the effective number of posterior draws, which should be at least 10 times the number of chains—40 in this case (Gelman et al., 2013, p. 287). Moreover, the $\hat{R}$ statistic needs to be below 1.1 for all variables, to ensure that a sufficient number of draws have been discarded as warm-up, and that the sampling is indeed performed from the actual posterior (Gelman et al., 2013, p. 287). For each model and variable of interest I check that these values are in the desired ranges, and they are reported in Table 6 for the main specification. All inference is performed using the Stan implementation in R (Carpenter et al., 2016; Gelman et al., 2013, Ap. C).

Table 5 shows the Bayesian hierarchical equivalent of Table 4 to compare the different model specifications. The parameters shown are the means of the hierarchical priors—for example, the values of the $\mu_j$’s in Equation (8). The values in the table are the posterior means, which asymptotically correspond with the ML or OLS point estimates in classical inference. The figures in brackets are the posterior SDs, which asymptotically correspond with the standard errors in classical inference. Finally, the AIC is replaced with the WAIC (Watanabe–Akaike or widely applicable information

---

8Because Bayesian simulation starts from some random initial values, the first simulation draws will reflect the initial starting point, rather than the actual posterior distribution. Therefore, the first draws (called the “warm-up” or “burn-in”) need to be discarded until they can be expected to reflect the actual posterior (Gelman et al., 2013, p. 282).
criterion), as this better accounts for the effective number of parameters used in hierarchical models, and is, in general, preferred in Bayesian analyses (Gelman et al., 2013, p. 173, 2014).

As was the case for the ML estimates, also in the Bayesian hierarchical model, the MA and second-order AR components are highly significant, and the WAIC markedly decreases when these components are added. Both suggest that the ARMA(2,2) specification best fits the data.

Table 6 shows more details for the ARMA(2,2) specifications for the main variables of interest: the mean of the resource advantage persistence rates $\lambda^{(k)}$, as well as the SD across industries. Note that the posterior SD in brackets is quite different from the SD of the $\lambda^{(k)}$ parameters in the bottom two rows: the SDs in brackets refer to the variation in posterior samples and are analogous to the classical standard errors, while the bottom rows with hierarchical SDs have no direct classical analogue, and refer to the variation of parameters across industries, thus providing an estimate to what extent industries are similar or different in terms of persistence patterns.

All variables are significantly larger than zero, and the posterior estimate for the mean of $\lambda^{(1)}$ is significantly higher than the 0.7–0.9 range common in the literature. The variation across industries can be seen from the relatively large SD priors, especially for $\lambda^{(2)}$. Assessing the full distribution of all posterior draws of $\lambda^{(k)}_j$ across industries indicates a positive $\lambda^{(2)}_j$ for the vast majority (99.3%) of posterior draws across industries—and thus the importance of the AR(2) component of the model.

9Strictly speaking, significance testing is not possible in Bayesian analysis, and instead one should speak of “posterior probability”. Specifically, I use the term “significant variable”, to signify a variable that has a high posterior probability to be strictly greater (or strictly smaller) than zero. Asymptotically, this is consistent with the use of the term “statistical significance” in classical inference.

### Table 5  Model comparison for Bayesian inference

<table>
<thead>
<tr>
<th></th>
<th>AR(1)</th>
<th>ARMA(1,1)</th>
<th>ARMA(2,1)</th>
<th>ARMA(2,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $\lambda^{(1)}$</td>
<td>0.866</td>
<td>0.886</td>
<td>0.949</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Mean $\lambda^{(2)}$</td>
<td></td>
<td></td>
<td>0.560</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Mean $\theta_1$</td>
<td>−0.052</td>
<td>−0.696</td>
<td>−0.648</td>
<td>−0.053</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.036)</td>
<td>(0.033)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Mean $\theta_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIC</td>
<td>182,279.8</td>
<td>181,772.2</td>
<td>181,490.2</td>
<td>181,296.5</td>
</tr>
</tbody>
</table>

Note. Bayesian posterior means for various hierarchical autoregressive moving average (ARMA) models of company excess profits. Posterior standard deviations in brackets (asymptotically, these correspond to standard errors in classical inference). The Watanabe–Akaike or widely applicable information criterion (WAIC) provides a Bayesian estimate of model fit: lower values indicate a better fit to the data, accounting for the effective number of parameters in the model.

### Table 6  Posterior inference for ARMA(2,2) model

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>2.5%</th>
<th>97.5%</th>
<th>n$_{\text{eff}}$</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $\lambda^{(1)}$</td>
<td>0.954 (0.009)</td>
<td>0.937</td>
<td>0.972</td>
<td>427</td>
<td>1.00</td>
</tr>
<tr>
<td>Mean $\lambda^{(2)}$</td>
<td>0.528 (0.031)</td>
<td>0.466</td>
<td>0.587</td>
<td>431</td>
<td>1.00</td>
</tr>
<tr>
<td>$SD \lambda^{(1)}$</td>
<td>0.048 (0.007)</td>
<td>0.036</td>
<td>0.064</td>
<td>500</td>
<td>1.00</td>
</tr>
<tr>
<td>$SD \lambda^{(2)}$</td>
<td>0.122 (0.016)</td>
<td>0.091</td>
<td>0.152</td>
<td>500</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. Posterior mean, standard deviation, and 95% interval bounds for the eigenvalues of the autoregressive components in the hierarchical ARMA(2,2) model of excess profits. The effective sample size $n_{\text{eff}}$ and the scale reduction factor $\hat{R}$ are used to assess convergence of the posterior draws; Gelman et al. (2013) recommend these to be above 40 and below 1.1, respectively.
signifying that higher-order resources significantly affect profit persistence patterns in (almost) all industries.

Appendix C of Supporting Information Appendix S1 includes several robustness checks to the analyses in this section.

4.3 | Implications for duration of competitive advantage

The results in Tables 4–5, and 6 clearly establish that the ARMA(2,2) model provides a better fit than the classical AR(1) model; that the AR(1) coefficient with this new specification is much larger than in the classical model; and that the AR(2) component is statistically highly significant. However, by themselves these results do not directly establish the economic significance of differences across these estimates.

To assess the economic implications of the different estimates of the λ-parameters, I run simulations of the resource evolution for the classical AR(1) estimate from Table 4 compared to the Bayesian ARMA(2,2) estimate from Table 6. To assess the effect of accounting for higher-order resource specifically, for the ARMA(2,2) model estimates I simulate once with only the λ₁ parameter, and another time with both λ-parameters. For each parameter set I run an AR(2) time-series simulation with normally distributed error terms for 1,000,000 time intervals—MA terms are omitted because these represent annual noise terms, while the underlying resource (dis-) advantage is of primary interest. The number of time intervals divided by the number of times each simulated time series crosses zero provides an estimate of the average duration of competitive advantage. Thus, competitive advantage is defined as having a resource position that provides a better-than-average expected value of profit (i.e., E_y > 0).

Figure 2 displays the results. Using only the λ₁-parameter, the average duration of competitive advantage in the classical AR(1) estimate (column a) is 5 years, about half the duration from the new estimate using the Bayesian ARMA(2,2) model (column b). Adding the effect of higher-order resources through λ₂ almost doubles the average duration of competitive advantage again to approximately 18 years. Thus, the models presented in this paper suggest a much longer duration of competitive advantage than usually assumed in the literature.

4.4 | Industry analysis

An attractive feature of hierarchical Bayesian modeling is that, in addition to inference at the level of the entire sample, it allows inference at the industry level.10 Figure 3 shows the mean and 95% interval of the λ-parameters, averaged by industry sector.11 Consistent with the results in the previous section, all 95% intervals are well above 0, signifying the importance of both AR components. Moreover, the results indicate clear differences across industries. Materials, Energy, and Utilities have relatively low values for both λ-parameters, suggesting lower persistence of profit differences in these industries—notably the low λ⁽¹⁾ value for utilities stands out. In these industries industry-level factors, such as energy prices, play an important role in profitability patterns, presumably diminishing the importance of both operating and higher-order resources, and thus profit persistence. On the other hand, Health Care, Industrials and Consumer Staples have relatively high values for both λ-parameters, with the value of λ⁽¹⁾ close to 1, suggesting a strong persistence of any profitability differences. Indeed, these industries are often associated with strong firm-specific resources (e.g., brands for

10Thanks to the editor Brian Wu and two anonymous reviewers for suggesting this addition to the paper.
11The figure shows the average by two-digit GISC sector of the original six-digit GICS industry estimates, weighted by the inverse of the posterior variance for each estimate (which is the optimal weighting).
FIGURE 2  Effect of $\lambda$-parameters on duration of competitive advantage. Note. Average length of competitive advantage duration based on simulations of resource evolution using (a) the classical AR(1) estimate for $\lambda_1$ from Table 4, (b) the $\lambda_1$ estimate from the ARMA(2,2) model in Table 6, and (c) both $\lambda$-parameters from the Bayesian ARMA(2,2) model.

FIGURE 3  Inference of $\lambda$-parameters by industry. Note. Weighted average of posterior industry means (dots) and 95% intervals (error bars) of $\lambda^{(1)}$ and $\lambda^{(2)}$ by two-digit GICS industry sector, sorted by increasing $\lambda^{(1)}$. 

<table>
<thead>
<tr>
<th>Industry Sector</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Health Care</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Information Technology</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Consumer Discretionary</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Consumer Staples</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Materials</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Telecommunication Services</td>
<td>0.954</td>
<td>0.528</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.954</td>
<td>0.528</td>
</tr>
</tbody>
</table>
consumer goods companies), which according to the RBV should lead to higher persistence of profits, consistent with the findings in, for instance, Villalonga (2004).
Additionally, it is interesting to analyze how industry-level patterns evolve over time. Therefore, I performed a Bayesian inference of the ARMA(2,2) model for each of the three decades between 1985 and 2014. Figure 4 shows the posterior mean and 95% intervals for $\lambda^{(1)}$ and $\lambda^{(2)}$, both for the average across industries (panel (a)), as well as for selected individual industries with particularly low or high values in some of the decades for either of the $\lambda$-parameters (panels (b)–(f)).

A clear time pattern stands out from this figure: $\lambda^{(2)}$ is lower in the decade 1995–2004 than in the decades before and after. This is consistent with both the findings of “hypercompetition” literature in the 2000s (e.g., D’Aveni et al., 2010), as well as recent findings of a reversion of this trend (Bennett & Gartenberg, 2016). Interestingly, the decline in persistence of profits in the decade 1995–2004 is almost exclusively in $\lambda^{(2)}$ (the persistence of profit growth), rather than in $\lambda^{(1)}$ (the persistence of profit levels). For some industries, $\lambda^{(2)}$ in the decade 1995–2004 is actually not significantly different from zero.

Another interesting finding is that some industries had periods with a $\lambda^{(1)} \geq 1$, such as Tobacco in 1985–1994 ($p = 0.02$). This means that during this decade any differences in profits were, on average, amplified over time. Note that this would have been troublesome had I used a traditional MLE, because the model would have been nonstationary (e.g., Hamilton, 1994). The issue is that the time-series models used in the MLE are based on the assumption that the time series can be extended ad infinitum. Indeed, this would be troublesome as any differences in profits would eventually grow arbitrarily large, which is impossible. Conceptually there is no problem with a $\lambda \geq 1$ though for a finite period of time; indeed it appears quite plausible that such “non-stationary episodes” could exist. In a Bayesian model it is rather easy to alleviate the assumptions that cause the trouble in the MLE, and thus provide correct inference for data with such non-stationary episodes.13

5 | DISCUSSION AND CONCLUSION

Evidence from the autocorrelation analysis, ML estimates, Bayesian analysis and the robustness tests all provide support for (a) the existence of at least one MA term and a resulting higher AR(1) component than found in preceding studies of profit persistence, sometimes even above unity, leading to “non-stationary episodes”, (b) the existence of a AR(2) component in profitability persistence patterns, and (c) a significant variation across industries in the persistence of operating and higher-order resources.

Based on the model used here, all these findings have a direct interpretation in the RBV framework. First, the higher AR(1) parameter suggests that profits are more persistent than found hitherto, further strengthening the importance of the RBV for strategy. Second, the existence of a higher-order AR component suggests the importance of higher-order resources in shaping profitability patterns. Third, the industry variation suggests that differences across industries are not only manifest in levels of profits, but also in the temporal dynamics of profits, and that resource-based theories can inform the differences in these dynamics.

In addition to these empirical findings and their theoretical interpretation, the paper shows how Bayesian hierarchical analysis—a method that is increasingly common in other disciplines, but still little used in strategy—can provide new insights on core strategic questions. The hierarchical model provides full statistical power across industries, without having to make the assumption that

12 Thanks to an anonymous reviewer for investigating potentially nonstationary behavior.

13 Using simulated data, I confirm that the Bayesian model used here provides the correct inference for nonstationary ARMA models, while the classical MLE fails. Details are available from the author upon request.
competitive dynamics are the same across industries. Moreover, the Bayesian framework allows making inferences at different levels of the hierarchy: both at the aggregate level and at the level of the industry—in principle, this could even be extended to the level of the individual company (Mackey, Barney, & Dotson, 2017). The Bayesian inference also provides more flexibility in, for instance, the estimation of nonstationary models, with AR components above unity.

Moreover, the theoretical model provides a different perspective on the relation between managerial agency and random variation. These two are usually viewed as mutually exclusive: the random variation is the null hypothesis, and managerial agency is the alternative explanation if the null hypothesis is rejected (e.g., Denrell, Fang, & Liu, 2014). Instead, in this paper, managerial agency can be endogenized in the stochastic model, through the variation in higher-order resources $z_t$. In this view, the quality of managers is itself a distribution that yields shocks to the stock of a firm's higher-order resources ($w_t$ in the model). In other words, from the standpoint of the researcher, a particularly skilled managerial team can be seen as a favorable draw from a distribution, providing higher-order resources that allow firms to obtain favorable operating resources, and generating profits at a fraction of their economic costs. This view is perfectly consistent with the standpoint of the manager that she or he can influence the company (and the broader world) beyond mere random variation.

The findings here are of direct relevance for managers. Most importantly, the much longer duration of competitive advantage implied by the model (Figure 2) underlines the importance for managers to analyze their firms’ resource positions vis-à-vis competitors, and to build strategies on their unique resource strengths. Moreover, managers should not only take into account operating resources such as brands, patents, captive customers, or specialized plants, but also higher-order resources such as strategic planning, M&A capabilities, and superior forecasts, as these can have a major impact in the long term.

Naturally, the study has several limitations, which could be addressed in future research. First, it is based on a single data source of accounting data for U.S. firms; this could be extended to other geographies as well as other types of data—in fact a Bayesian modeling structure would be uniquely suited to incorporate data from multiple sources (e.g., accounting and stock market data) in a single model. Another improvement would be to consider a broader class of models, for instance including interactions between operating and higher-order resources, the addition of nonlinear terms, and models that would allow for changing error term distributions over time—for example, as a firm gets bigger it is highly plausible that profitability shocks in absolute terms also get larger, thus it would be helpful to incorporate that into the model. Finally, it would be interesting to further explore the differences in resource dynamics resulting from business unit versus corporate versus industry effects, and investigate their origins—for instance to explain drivers of the differences in profit persistence across industries and over time.

ACKNOWLEDGEMENTS

I wish to thank Emilie Feldman, David Hsu, Shane Jensen, Dan Levinthal, Katherine Klein, Anoop Menon, Arkadiy Sakhartov, Paul Shaman, Nicolaj Siggelkow, and Bart Vanneste, as well as session participants at the SMS annual conference 2014 in Madrid, the Wharton-INSEAD doctoral alliance meetings in 2014 and 2015, the Transatlantic Doctoral Conference 2016 in London, the Tilburg Organization Studies seminar, the Academy of Management conference 2016 in Anaheim, the Strategic Organization Design seminar at the University of Southern Denmark, and the Wharton PhD seminar for valuable feedback. Moreover, I want to thank the editor Brian Wu and two anonymous reviewers for a particularly constructive review process and for their excellent feedback, which has significantly improved the paper. Finally, I want to thank the staff at the Wharton High Performance
Computing Cluster (HPCC) for technical support and Hamelin Hazel for editorial support. Any errors remain my own.

ORCID

Phedo D. Wibbens https://orcid.org/0000-0002-1517-3858

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Wibbens PD. Performance persistence in the presence of higher-order resources. *Strat Mgmt J*. 2019;40:181–202. [https://doi.org/10.1002/smj.2979](https://doi.org/10.1002/smj.2979)