Central clearing: Risks and customer protections

Ivana Ruffini

Introduction and summary
Central clearing shifts risk, sometimes reduces it, but does not eliminate it. In the wake of the 2008 global financial crisis, the Group of Twenty (G-20) developed a regulatory reform program for derivatives contracts, with a stated goal of reducing systemic risk by requiring a market structure shift from a bilateral framework to a centrally cleared framework for standardized over-the-counter (OTC) derivatives (Group of Twenty, 2009).

OTC bilateral transactions are usually collateralized directly between the counterparties, while central clearing generally involves the use of one or more intermediaries in the clearing and settlement process. Under a bilateral framework, the exposures that participants face can be dispersed across a large number of counterparties, while under the centrally cleared framework these risks are shifted to and concentrated in central counterparties (CCPs) and financial intermediaries, such as clearing members (CMs) and futures commission merchants (FCMs).

In the United States, since 1936 the segregation of customer funds from intermediaries’ house funds has been the key mechanism for customer protections in intermediated derivatives markets. After the 2008 financial crisis, as part of the overhaul of the financial regulatory system, regulators enacted rules aimed at improving systemwide management of counterparty risk. As a result, new customer protection frameworks and requirements for central clearing of standardized swaps were implemented. Although the lack of harmonization of these new rules between different jurisdictions may introduce additional complexities, the primary focus of this article is on the centrally cleared markets that fall under the regulatory authority of the U.S. Commodity Futures Trading Commission (CFTC).

This article examines the impact of the market structure change and associated customer protection frameworks on risks faced by market participants, with a focus on the liquidity and credit risks that could arise in the aftermath of a potential FCM failure.

The article is organized into four sections. First, I provide a brief overview of central counterparty clearing. Second, I describe the salient characteristics of intermediation in centrally cleared markets. Third, I define the key risks associated with the failure of a clearing member and provide examples to illustrate the variability of exposures. Finally, I discuss customer protection frameworks that are used to mitigate the impact of the identified exposures.

Ivana Ruffini is a senior policy specialist on the financial markets team in the Economic Research Department at the Federal Reserve Bank of Chicago. The author would like to thank David Marshall, Robert Steigerwald, John McPartland, and Robert Cox for thoughtful comments and insights.

© 2016 Federal Reserve Bank of Chicago

Economic Perspectives is published by the Economic Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors’ and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Charles L. Evans, President; Daniel G. Sullivan, Executive Vice President and Director of Research; David Marshall, Senior Vice President and Associate Director of Research; Spencer Krane, Senior Vice President and Senior Research Advisor; Daniel Aaronson, Vice President, microeconomic policy research; Jonas D. M. Fisher, Vice President, macroeconomic policy research; Robert Cox, Vice President, markets team; Anna L. Paulson, Vice President, finance team; William A. Testa, Vice President, regional programs; Lisa Barrow, Senior Economist and Economics Editor; Helen Koshy and Han Y. Choi, Editors; Julia Baker, Production Editor; Sheila A. Mangler, Editorial Assistant.

Economic Perspectives articles may be reproduced in whole or in part, provided the articles are not reproduced or distributed for commercial gain and provided the source is appropriately credited. Prior written permission must be obtained for any other reproduction, distribution, republication, or creation of derivative works of Economic Perspectives articles. To request permission, please contact Helen Koshy, senior editor, at 312-322-5830 or email Helen.Koshy@chicagofed.org.

ISSN 0164-0682
Central counterparty clearing

Central counterparty clearing refers to the post-trade process of counterparty substitution, whereby a single counterparty (clearinghouse) replaces the original counterparties in all centrally cleared contracts and the clearinghouse becomes the sole counterparty to all CMs. This counterparty substitution results in an exchange of the credit risk exposure of the “original counterparties for the credit risk of the CCP” (Culp, 2010, p. 10).

Central counterparty arrangements for exchange-traded contracts evolved organically “when gains from the intermediated exchange exceed[ed] the gains from the direct exchange” (Spulber, 1999, p. xiii). Originally, CCPs were established by the CMs to facilitate clearing and settlements of trades and until fairly recently, CCPs were owned by their CMs. This mutualization required all CMs to comply with risk controls to limit the extent to which the trading activities of any individual CM could expose other CMs to potential losses. CCPs restricted membership in the clearinghouse to those institutions that could comply with strict membership and risk-management criteria. Over the years, many CCPs have demutualized and become part of publicly traded companies. CCPs continue to enforce strict membership and risk-management standards and require CMs to contribute to the CCP guarantee fund (Murphy, 2013, p. 214).

In this section, I explain how central clearing arrangements benefit CMs and end-users through multilateral netting, collateralization of positions, transparent pricing, and default management. Nevertheless, these benefits can be costly and some contend that “the fact that not all OTC derivatives have flooded into a CCP is a strong indication that there are both costs and benefits associated with central clearing” (Culp, 2010, p. 15).

Multilateral netting allows for the aggregate offset of positions and the termination of economically redundant obligations. Multilateral netting offsets obligations between multiple parties as opposed to bilateral netting, which offsets obligations between only two counterparties. A shift in counterparty exposures to a centralized structure allows for this multilateral netting of obligations, often resulting in a reduction of counterparty credit risk and the liquidity risk borne by CMs. Figure 1 illustrates the mechanism of counterparty substitution and the impact of multilateral netting.

The change in the exposure between bilaterally and centrally cleared trades can be significant. Figure 1 shows that the exposure of $270 is reduced to $40 as a result of multilateral netting. In this example, the multilateral netting reduces counterparty credit and liquidity risk exposures by replacing the bilateral obligations between counterparties with a new obligation between each clearing member and the CCP. However, multilateral netting may not always be more efficient than bilateral netting. On a global scale, central clearing is fragmented across legal jurisdictions and as a result of such fragmentation, multilateral netting can sometimes actually increase the expected exposures compared with bilateral netting arrangements. For example, Duffie and Zhu present a case involving credit default swaps (CDS), in which “clearing the U.S. and European CDS separately increases expected exposures by 9% relative to bilateral netting” (Duffie and Zhu, 2011, p. 87).

Collateralization of positions refers to the practice of posting collateral to the counterparty in a derivatives transaction to ensure compliance with the counterparty margin requirements. Margining of positions collateralizes the risk exposure of the CCP to CMs and of CMs to market participants. However,
those who post margin may face liquidity and credit risk exposure, as I discuss in more detail later.

There are different types of margin requirements—some can be satisfied with securities and others only with cash. For example, in centrally cleared markets, market participants must deposit collateral (initial margin) with the FCM to open a margin account and participate in the marketplace. FCMs keep their own funds in a “house” account and are required to keep customer margin in the customer segregated account. Also, they are required to extend the segregation framework to the CCP in the way they transfer customer margin assets with the CCP.

CCPs set minimum initial margin requirements. CMs guarantee their clients’ positions to the CCP and may require their clients to post more collateral than the CCP requires. Initial margin is required for all open derivatives positions and reflects the margin period of risk, the CCP’s best estimate of the number of days that it would likely take the CCP to liquidate or auction a portfolio of positions. Variation margin is the periodic mark to market of positions that effectively restores margin to its original level. In this way, a CCP can operate prudently with initial margin levels that only reflect a reasonable margin period of risk. Variation margin is always paid with cash.

**Transparent valuation** of margined assets and positions is a feature of centralized clearing that limits “disputes about collateral valuation” (Culp, 2010, p. 16) and thus reduces the likelihood of procyclical liquidity shocks, such as those observed during the 2008 global crisis in OTC CDS markets. Compared with the bilateral arrangements in which collateral requirements and valuations can vary from counterparty to counterparty, CCPs have a common approach to collateralization and valuation that is consistent across all CMs. The CCP rulebooks are public documents that specify rules of conduct and consequences that follow certain actions or changes in exposures. Furthermore, CCPs also communicate methodologies for the calculation of margin requirements and settlement obligations with their CMs. Additionally, the consistency with which CCPs apply the rules across all CMs further eliminates uncertainty about the value of collateral pledged to support cleared positions, facilitating CM management of liquidity risk exposure.
Default management, loss allocation, and default waterfall are specified in the CCP rulebooks and facilitate orderly management of CM defaults. CCPs set aside some of their own capital to cover a portion of a loss incurred by the CCP as the result of a CM default. CCPs also collect guarantee fund contributions from each CM to fund their mutualized guarantee pool, generally commensurate to the risk that individual CMs pose to the CCP. Regulatory requirements set the minimum standards for determination of the guarantee fund size. The CCP is responsible for the variation margin obligations of the defaulter’s positions until those positions have been liquidated or assumed by a solvent CM. Any potential financial loss associated with doing so would initially be covered by liquidation of the defaulter’s margin deposits and the defaulter’s contribution to the CCP’s guarantee fund. If the losses were to exceed the value of the defaulter’s assets at the CCP, the remaining loss would be absorbed by a combination of the CCP’s capital and guarantee fund, which includes the contributions of the nondefaulting CMs.

While the surviving CMs may have an indirect exposure to a failed CM, any customer margin assets of the surviving CMs are not involved in the default process and thus are protected from such indirect exposures. However, the customer assets and positions of the defaulted CM are not protected, as the CCP stands only as counterparty to the CM (the financial intermediary). A CM’s client assets are not exposed to this default risk unless the default occurs in the customer origin. CCPs only guarantee the performance of CMs to the other CMs.

Financial intermediation

Financial intermediaries are an integral part of the clearing structure. Intermediation helps CCPs manage their counterparty risk exposure by limiting direct access to the clearinghouse to its members. Membership criteria are demanding, and many market participants don’t qualify to become CMs. For that reason, many CMs serve as financial intermediaries to market participants.

All CMs must contribute to the CCP guarantee fund and comply with various regulatory, capital, risk-management, and operational requirements. Additionally, CMs must agree to guarantee and assume responsibility for all trades that they submit for clearing (CME Group, 2015). It is important to highlight that counterparties to a centrally cleared transaction are only a CCP and a CM—market participants that are not CMs have no direct claim upon the CCP. In other words, a CCP only guarantees that it will honor its contractual obligations to its CMs.

In the U.S. derivatives and futures markets, trade intermediaries that handle customer assets must be registered with the CFTC as futures commission merchants (FCMs). They may serve as brokers, custodians, and guarantors for their clients’ transactions. FCMs do not have to be clearing members; and when they are not, they require another layer of intermediation—FCM (D) in figure 2 is an example of an FCM that is not a clearing member. FCMs hold customer assets and margin collateral in commingled customer segregated omnibus accounts as depicted in figure 2. CFTC rules permit operational commingling of customer assets through an omnibus account structure. In general, CFTC rules prohibit the use of the margin assets of one client to offset a potential margin deficiency (or any obligation) of another client in a customer segregated account. Still, the intermediation and pooling of all customer assets/collateral in one account can expose the non-defaulting customers to potential losses in the event that fellow customers and the FCM fail and the aggregate customer margin assets fall short of the total claims of customers on the failed FCM’s pool of customer segregated assets (Culp, 2013).

FCMs that clear trades for themselves and their customers have a house account for their own trades and a customer segregated account for their customers. The blue dotted line in figure 2 represents the flow of transactions submitted for central clearing, while the red and green lines represent customer and house payment flows, respectively.

FCMs routinely extend intraday credit to their clients, because FCMs are typically required to complete settlements with the CCP before they settle with their individual clients. Most customer accounts are not prefunded. FCMs transfer house funds to supplement any potential shortages in their customer segregated accounts. This practice is encouraged by the regulators (Futures Industry Association, 2013). Once the required customer margin payments are received (and any deficiencies covered), the FCM returns their funds to their house account. FCMs often simply maintain a surplus of house funds in their customer segregated funds accounts for ease of operation, known as their residual interest.

The centrally cleared market structure does not eliminate counterparty risk. The market structure change does not just concentrate the counterparty risk in the CCP, but it also introduces new counterparty risk exposures. The FCM intermediaries introduce some new exposures to both fellow customers and the FCM itself. Quantifying such exposures is complicated and somewhat obscured by the different levels of intermediation inherent in the centrally cleared market structure.
Customer risks

CMs guarantee all matched trades that are submitted for clearing and act as a secured custodian over client margin assets to ensure financial performance of their customers in the aggregate. When a financial intermediary (FCM) fails, the customers are exposed to liquidity and credit risk. Customers may incur losses due to a delay in immediate availability of funds/assets or a direct loss of money/assets.

Liquidity risk

Liquidity is the ability to fund, satisfy commitments in a timely manner, and transact in financial markets without suffering severe losses (BIS, 2004). In general, uncertainty about the financial health of counterparties has a negative effect on their liquidity (Afonso, Kovner, and Schoar, 2010; Heider, Hoerova, and Holthausen, 2009; Freixas and Jorge, 2008; and Flannery, 1996). Liquidity risk can propagate and magnify market and counterparty credit risks, thereby spreading liquidity shocks throughout the financial system. Most central banks serve as “lenders of last resort” in order to curb this propagation of liquidity shortages and foster financial stability (Cecchetti and Disyatat, 2010). There are many different types of liquidity, but two that best capture customer exposure to liquidity risks inherent in intermediated derivatives markets are market liquidity and funding liquidity.

Market liquidity refers to the market’s capacity for trading large quantities of assets without an uncharacteristic price impact or “the ease with which an asset can be converted into means of payment” (Cecchetti and Disyatat, 2010, p. 30). The 2008 financial crisis is a perfect example of the impact of market illiquidity. At times during the crisis, there was no market at all for certain assets such as mortgage-backed securities (MBS). This lack of a market at any price led to uncertainty. Market participants that held MBS experienced distress as it became increasingly difficult to accurately revalue their MBS holdings without a functioning secondary market.

It is important to realize that in centrally cleared markets, a decline in the market liquidity of a particular asset class can precipitate a decline in the post-haircut value of margin collateral for market participants that have pledged such assets. The FCM

---

**FIGURE 2**

Simplified centrally cleared market structure

Source: Federal Reserve Bank of Chicago.
intermediaries would likely request additional collateral from affected clients to provide the additional margin as required by the CCP.

Another type of liquidity risk exposure that is magnified in centrally cleared markets is the lack of funding liquidity. Funding liquidity refers to the existence of abundant and diverse sources of cash for market participants. One example of funding liquidity is “just-in-time” liquidity, which represents the ability of market participants to make payments that specify location, currency, and “a precise time frame measured not in days, but in hours or even minutes” (Heckinger, Marshall, and Steigerwald, 2009). Trading in futures and cleared swaps markets involves the use of funds necessary to satisfy margin requirements (Brunnermeier and Pedersen, 2009). Margin collateral is valued multiple times during the day, and any shortage is required to be funded as part of the next clearing cycle.

Market and funding liquidity exposures are not mutually exclusive. Changes in market liquidity can negatively impact the value of margin collateral and put pressure on financial intermediaries to provide funding. This, in turn, can negatively impact funding liquidity and result in broader uncertainty in the marketplace. Uncertainty can cause a disruption in just-in-time liquidity, as market participants take extra time to evaluate their contractual obligations, leading to a systemic shortage of liquidity:

A systemic shortage of both funding and market liquidity … is potentially the most destructive. It involves tensions emanating from an evaporation of confidence and from coordination failures among market participants that lead to a breakdown of key financial markets. (Cecchetti and Disyatat, 2010, p. 31)

In centrally cleared markets, the customers of a failed FCM face uncertainty with respect to their ability to transfer trades and margin to another (solvent) FCM. Also, customers of failed FCMs may face direct losses due to a shortage in the value of the aggregate pool of customer segregated assets.

FCMs are also permitted to invest customer funds. If such investments suffered a decline in value, and if at the same time the FCM failed, it is conceivable that customers could incur a loss. Furthermore, in times of liquidity stress, if an FCM fails, the customers’ margin assets might not necessarily be immediately accessible, and those clients themselves could default or even become insolvent as a result.

**Credit risk**

Prior to the failure of MF Global and Peregrine Financial Group (PFG), many underestimated the risks associated with intermediation in centrally cleared markets. Some assumed that customer segregation meant that their funds were segregated both from other customers and from the FCM house account. In financial markets, commingling serves valuable purposes of streamlining operations, funding the business, and reducing day-to-day costs for customers. However, such benefits come at a cost of exposure to risks, primarily in the form of credit exposure to financial intermediaries and to fellow customers of such intermediaries.

Failures of intermediaries are quite rare, and customer losses resulting from such failures are rarer still. Historically, inadequate management of operational exposures, including fraud, has been the primary cause of many FCM failures.

As an example, in 1995 Barings Bank failed when a rogue trader accumulated substantial proprietary trading losses. Global futures customers did not suffer loss of margin as Barings Bank was purchased by the Dutch bank, ING, which assumed all of Baring’s liabilities (Bank of England, 1995).

In contrast, a more recent failure of an FCM, PFG, did result in substantial customer losses. Fraudulent behavior was uncovered in 2012. For several years, PFG management had been fabricating audit confirmation replies that were sent to its regulator in order to conceal an ongoing embezzlement of customer segregated funds. According to the bankruptcy trustee, PFG had embezzled about $200 million of its customer segregated funds (Peterson, 2014). PFG was not a CM of any CCP and thus not subject to the audit regimen of a major CCP.

Customers can also be exposed to losses due to the failure of other customers of the same financial intermediary FCM. Historically, such losses have been so uncommon and so small that sometimes exchanges have opted to make clients of failed FCMs whole, even though they were not contractually obligated to do so: The Commodity Exchange in New York said Monday that it plans to advance $4.1 million to ensure that customers of the failed Volume Investors Corp. receive the money owed them. … The repayment plan had been a face-saving move for the Comex, which faced a barrage of industry criticism following the failure last March of Volume Investors, a Comex member. The incident was the first time customers stood to lose money because of the demise of a member of a futures exchange. (Cohen, 1985, p. 1)

In other instances, customers did lose money as a consequence of the failure of another customer. The case of Griffin Trading is one example. Griffin Trading filed for bankruptcy in 1998, because John Ho Park,
one of Griffin’s European customers, “sustained trading losses … and neither Park nor Griffin Trading had enough capital to cover these obligations.” Griffin’s solvent European customers lost a portion of their margin assets because Griffin management used funds in the omnibus customer segregated account of Griffin’s UK FCM to fund a margin call on Park’s trades.

The failure of MF Global illustrates a different problem. In the MF Global case, customers suffered losses because MF Global mishandled customer segregated funds. Customers of MF Global who waited until the end of the resolution of the estate actually received all of their funds back. Still, many other customers realized losses because they sold “their claims on MF Global to hedge funds and banks for roughly 90 percent or more of face value” (Protest, 2014, p. 1).

Fraud and a lack of operational robustness can expose FCM clients to considerable risk. Such risk can be realized as a loss in cases where there is a shortfall in a customer segregated account even if customer assets are held in an appropriate account. Surviving customers may incur losses not only due to a delay in the return of margin assets but also face the risk that the assets may not be recovered in full. In the United States, regulations and policies designed to protect customer margin assets are based on the segregation of such assets from the proprietary assets of the financial intermediary. However, these regulations can be constrained by countervailing provisions of §766(h) of the U.S. Bankruptcy Code. In cases of undersegregation of customer funds, the U.S. Bankruptcy Code would treat all surviving customers as the same class, regardless of whether their assets are in an omnibus account structure or individually segregated. Consequently, the surviving customers would share in the shortfall in the segregated funds on a pro rata basis (Futures Industry Association, 2012). To further limit such risk exposures, the CFTC enforces two customer protection frameworks.

**Customer protections**

In the United States, exposure to FCM risk is somewhat mitigated by the regulation of market intermediaries and the implementation of two customer protection frameworks. The traditional U.S. futures segregation framework applies to futures markets. The legally segregated operationally commingled (LSOC) framework applies to centrally cleared swaps markets. These frameworks rely on rules that govern segregation of customer assets held by intermediaries and CCPs.

**U.S. futures segregation model**

Segregation requirements for customer margin assets in U.S. futures markets are largely set out in section 4d(a)(2) of the Commodity Exchange Act and CFTC regulation 1.20. The section states that in a case of an FCM insolvency, the customer segregated funds at depository institutions are protected from the “banker’s right of setoff.” This would remove customer segregated funds deposited by an FCM or by a CCP from a bank’s right of setoff against any debts owed to that bank by that FCM or CCP.

In the case of an FCM bankruptcy, customer segregated funds are meant to repay customer claims. When the aggregate amount in customer segregated accounts equals what customers are owed, the customers are made whole. If there is an aggregate excess in the FCM’s customer segregated accounts, customers are again made whole and the excess (residual interest) margin that does not belong to customers is returned to the estate of the FCM. Conversely, if the aggregate pool of customer segregated assets is less than the aggregate claims of customers on the segregated pool, customers’ claims are distributed pro rata with all customers incurring the same percentage loss.

It is important to highlight that an undersegregation condition is a violation of CFTC rules and generally occurs due to fraudulent activity or operational problems (Culp, 2013). The U.S. futures segregation model does not attempt to address potentially fraudulent activity or operational failures. It is not designed to offer any additional protections to customers of insolvent FCMs with regard to the aforementioned risks.

**Legally segregated operationally commingled (LSOC)**

In the United States, segregation requirements for customer margin assets for cleared swaps markets are set out in section 4d(f) of CFTC regulations 22.2 and 1.22. LSOC is significant as it precludes the option of a CCP to utilize the initial margin assets of nondefaulting cleared swaps customers of a failed FCM to offset the financial loss of one or more defaulting cleared swaps customers of that FCM. It also differs from the traditional U.S. futures segregation framework in that it does attempt to reduce the risk of operational failures that might result in an undersegregated condition.

Under the LSOC framework, an FCM that clears swaps for customers is required to transmit account-level margin and position information to the CCP on a daily basis (U.S. Commodity Futures Trading Commission, 2012). Additionally, the CCP is required to validate and attest to the accuracy of that account-level information on a daily basis (U.S. Commodity Futures Trading Commission, 2012). These requirements significantly improve operational controls and
expand oversight over customer segregated funds to include the CCP. This practice reduces operational risks and reduces the potential for fraudulent behavior on the part of an FCM. Perhaps more importantly, LSOC has the potential to greatly facilitate the prompt and orderly transfer of the positions and the margin assets of the uninvolved cleared swaps customers of a failed FCM because the CCP would have the relevant account-level information in hand, before the fact. LSOC represents a departure from the traditional U.S. futures segregation model but remains constrained by §766(h) of the U.S. Bankruptcy Code.

LSOC “only explicitly protect[s] the collateral value attributed to each customer as reported by FCMs” (CME Group, 2012, p. 2) to CCPs. Additionally, any excess in customer margin deposited with the CCP for cleared swaps receives full protection under the LSOC framework in the case of an FCM default. Any excess margin would not be returned to the FCM’s estate, but would either be transferred together with the client positions to another FCM or returned to the swaps clearing market participant (U.S. Commodity Futures Trading Commission, 2012). Furthermore, any FCM residual interest in the cleared swaps customer segregated origin of the failed FCM at the CCP would be treated as customer segregated assets and would be protected under the LSOC framework (U.S. Commodity Futures Trading Commission, 2012).

Still, LSOC has its limitations. Section 766(h) of the U.S. Bankruptcy Code provides that “non-defaulting customers in an account class that has incurred a loss, e.g., the Customer Segregated Account, will share in any shortfall, pro rata” (Futures Industry Association, 2014, p. 9). An FCM’s customers remain exposed to potential pro rata losses should their FCM fail:

(i) if the bankrupt FCM’s books and records are inaccurate; (ii) in the event of a shortfall in the Cleared Swaps Customer Account arising from FCM fraud or mismanagement; or (iii) in the event a bankruptcy trustee incurs losses in liquidating collateral held in the Cleared Swaps Customer Account in which the FCM had invested in accordance with Commission Rule 1.25. (Futures Industry Association, 2014, p. 7)

Conclusion

During the 2008 financial crisis, uncertainty about the financial health of counterparties resulted in gridlock in the marketplace. Failure of bilateral counterparties to assess and address counterparty exposures increased systemic risk and had a negative impact on the broader economy. The central clearing mandate and LSOC were meant to reduce systemic risk, but it was not entirely eliminated.

Central clearing reduces risk through multilateral netting, collateralization of positions, pricing, and default management practices. However, central clearing also concentrates risk into a CCP, and financial intermediation introduces new risks. The concentration of risk in CCPs must not be underestimated, as CCP failures, while rare, do happen. Furthermore, while some failures of financial intermediaries, such as Lehman in 2008 and Refco in 2005, were successfully managed by central counterparties (Culp, 2010), other failures have resulted in customer losses. Such losses occurred when customer funds were misused by intermediaries—MF Global in 2011 and PFG in 2012; and when a customer defaulted—Griffin Trading in 1998.

Central clearing does not protect customers of a defaulting FCM. Customer protection frameworks are intended to mitigate such exposures. However, the protections offered under the traditional U.S. futures customer segregation and LSOC are somewhat limited. Both frameworks rely on segregation of customer funds to protect customer assets. However, under the U.S. Bankruptcy Code even individually segregated customer funds are treated as if they were held commingled in a single omnibus account.
NOTES

1Direct clearing members that clear their own trades don’t go through an intermediary.

2I use the acronym FCM here to mean a financial intermediary for customers that want to transact in centrally cleared markets, although this use is an oversimplification because FCMs are not necessarily direct clearing members.

3A complete list of CFTC registered participants and organizations, including derivatives clearing organizations (DCOs), derivatives contract markets (DCMs), and swap execution facilities (SEFs), is available on the CFTC’s website at http://www.cftc.gov.

4The Bank for International Settlements’ document, Principles for Financial Market Infrastructures (PFMI), outlines general requirements for guarantee fund calculations; however, regulations vary based on the interpretations of the PFMI by different regulatory authorities. It is available at http://www.bis.org/cpmi/publ/d101a.pdf.

5FCMs facilitate trade execution for their clients; they serve as custodians of customer property and are responsible for the collection and transfer of margin assets between customers and CCPs; and they guarantee the performance of their clients to the CCP. See http://www.nfa.futures.org/NFA-registration/fcm/index.HTML.

6Omnibus accounts are customer segregated accounts held at an FCM and include the commingled funds (cash, assets, and/or securities) of all customers of a particular FCM. The CFTC regulation 1.20 (17 CFR 1.20) states that all customer segregated funds are allowed to be placed in a single or omnibus account as long as the name of the account reflects that the funds are being held for the benefit of the CM’s customers. See http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/index.htm and http://www.ecfr.gov/cgi-bin/text-idx?c=ecfr&SID=9726fa13fed92e969b82107fdef0e6cf&rgn=div8&view=text&node=17:1.0.1.1.0.4.19&idno=17.


8Insurance solutions have been contemplated that can mitigate the losses that arise from the failure of a clearing member. Some insurance products are currently offered but have not been adopted by CCPs.


11Over the past 50 years, there have been several CCP failures associated with a market crisis—Paris, 1974; Kuala Lumpur, 1984; and Hong Kong, 1987 (Rehlon and Nixon, 2013).
REFERENCES


Auto production footprints: Comparing Europe and North America

Thomas H. Klier and James M. Rubenstein

Introduction and summary

Today’s footprints of motor vehicle production in Europe and North America appear at first glance to be remarkably similar: In both regions, plants producing motor vehicles are highly agglomerated, which is typical of manufacturing activities. The auto industry is a global industry: A dozen or so mass producers compete with one another around the world. Because these automakers employ similar production models in their plants, one might expect similar forces to shape their production location decisions. This article evaluates whether the same general factors explain the broad patterns seen in the auto industry’s footprints in Europe and North America. This question is of particular interest because to date, little comparative analysis of this kind has been performed, especially involving Europe as a whole. In general, most auto industry analysis of Europe has focused on its individual countries instead of the entirety of the region.

We begin the article with a description of the current distribution of motor vehicle production in both North America and Europe. Then we review the principles of agglomeration and industrial location theories and discuss their applicability to auto production siting decisions. Next, we examine whether these principles adequately explain changes in the geographical distribution of auto production in North America. We outline key events in Europe around 1990 that affected the spatial distribution of auto production there. And we evaluate to what extent the principles of agglomeration and industrial location theories are sufficient to explain the changing geography of auto production in Europe. In doing so, we also illustrate the growing importance of a northwest–southeast corridor in Europe, where the auto industry has become concentrated. Furthermore, we discuss trends in auto assembly plant openings and closings—both inside and outside this European corridor of production—since 1990. Finally, we highlight the features of auto production in Europe and North America that are not consistent with agglomeration theory.

The current geography of auto production in North America and Europe

Motor vehicle production involves two types of firms: vehicle assemblers and producers of parts (or parts suppliers). Today about a dozen carmakers put together light vehicles (see note 1) at approximately 80 assembly plants in Europe and approximately 70 assembly plants...
in North America. The roughly 15,000 parts that go into each vehicle are produced at several thousand parts supplier plants in both regions. For the purposes of this article, Europe is defined as the 16 member countries of the European Union (EU) that have produced at least 100,000 motor vehicles in any year between 1990 and 2013. The 16 countries are Austria, Belgium, Czechia, France, Germany, Hungary, Italy, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom (UK). In 2013, auto production reached at least 100,000 units in 15 of these 16 countries; the exception was the Netherlands, where auto production last hit 100,000 units in 2005. In this article, Central Europe refers to Czechia, Hungary, Poland, Romania, Slovakia, and Slovenia, while Western Europe refers to the other ten auto-producing countries. Here, North America refers to Canada, Mexico, and the United States.

Motor vehicle production in North America is clustered in a north–south corridor, mostly in the United States, called “auto alley” (see figure 1). This corridor is roughly 800 miles long and 250 miles wide, extending between Michigan and Alabama. The spine of auto alley is formed by the north–south interstate highways I-65 and I-75. Auto alley extends into Canada along Route 401 (Klier and Rubenstein, 2008; Klier and McMillen, 2006, 2008; and Rubenstein, 1992). Within the United States, auto alley accounted for nearly 90 percent of light vehicle production in 2013.

In Europe, motor vehicle production is clustered in a corridor along a northwest–southeast axis between the Danube River and the North Sea, with an extension across the English Channel into the United Kingdom (see figure 2). This corridor is roughly 800 miles long and 250 miles wide; it encompasses nearly the same amount of area and has almost the same shape as North America’s auto alley. In Europe, the corridor...
of motor vehicle production encompasses the assembly plants of the United Kingdom, northeastern France, Belgium, the Netherlands, Germany, Austria, southern Poland, Czechia, Slovakia, and Hungary. The corridor lies roughly along the major east–west highways E30 and E50. Its eastern and western ends are approximately equivalent to the maximum distance that truck drivers can reach in one day from southwestern Germany—Europe’s economic and population center.  

It is remarkable that the motor vehicle production corridors in North America and Europe do not just appear rather similar but also represent comparable shares of their respective regions’ total auto plants. Approximately 73 percent of North America’s auto assembly plants and 62 percent of its parts supplier plants are located in auto alley, including the Canadian extension. And approximately 73 percent of Europe’s vehicle assembly plants and 74 percent of its parts supplier plants are located in the auto production corridor, including the UK extension.

Most of North America’s motor vehicle production outside auto alley takes place in Mexico, which is home to 19 percent of the region’s assembly plants and 20 percent of its parts supplier plants. In Europe, Spain is the leading area of auto assembly outside the corridor, and Romania and Italy are the leading areas of auto parts production outside the corridor (Frigant and Miollan, 2014).

**Agglomeration and industrial location theories**

Agglomeration is the association of productive activities in proximity to one another (Gregory et al., 2009, p. 14). As shown in figure 3, three competitors may independently compute each of their optimal plant
sites as locations A, B, and C. But if all three locate at location X, they benefit from agglomeration economies. According to Marshall (1920), agglomeration can reduce the cost of obtaining inputs and shipping final goods, the cost of moving workers across employers, and the cost of disseminating new ideas (thereby encouraging “knowledge spillovers” and faster rates of innovation). According to Ellison, Glaeser, and Kerr (2010, p. 1195), “the benefits of agglomeration ultimately reflect gains that occur when proximity reduces transport costs,” which include not only shipping costs but also the costs of moving employees and ideas.

As noted before, manufacturing activity tends to be agglomerated (see, for example, Krugman, 1991; Ellison and Glaeser, 1997; Head and Mayer, 2004; and Duranton and Overman, 2005, 2008). Among manufacturing industries, the auto industry is consistently ranked as one of the most agglomerated (Ellison and Glaeser, 1997; Duranton and Overman, 2005; and Goldman, Klier, and Walstrum, 2015). Indeed, there tends to be a high degree of co-location of auto assembly plants and parts supplier plants. Competing carmakers have, in many cases, placed their assembly plants fairly close to one another geographically. Moreover, these assembly plants often share a network of parts suppliers that are within a reasonable distance from their locations. The proximity of auto assembly plants to parts suppliers can result in lower prices for inputs for carmakers; meanwhile, suppliers benefit from being able to do business with multiple auto assembly customers.

Industrial location theory helps researchers within the field of economic geography better understand and explain plant location decisions. The theory comes from the work of Alfred Weber (1929). Weber argued that the optimal location for a factory is the point that minimizes the aggregate costs of bringing in inputs from suppliers and shipping out final products to consumers. So, according to Weber, the least-cost location can be computed from a geometric model. As shown in figure 4, the optimal location for a factory with one market and two sources of inputs is a point that minimizes the aggregate cost of shipping the two inputs to the factory and shipping the finished product to the market.

In his theory of industrial location, Weber (1929) distinguishes between two types of industries—namely, bulk-reducing industries and bulk-gaining industries. In a bulk-reducing industry—one with inputs that are heavier or occupy a greater volume than the final product—production facilities tend to locate near the sources of inputs in order to minimize the shipping costs (see the schematic on the left in figure 4). Conversely, in a bulk-gaining industry—one whose fabricated product is heavier or occupies a greater volume than the inputs—production facilities have a tendency to be near the markets for the final good in order to minimize the shipping costs (see the schematic on the right in figure 4).

Copper and steel provide examples of bulk-reducing industries. For example, in the United States, most copper concentration mills, smelters, and refineries are located near copper mines in Arizona (Ó hUallacháin and Matthews, 1994; and Rubenstein, 2014). In addition, in the United States, integrated steel mills are clustered in the southern Great Lakes region to minimize the aggregate shipping costs of the two principal inputs—coal from Appalachia and iron ore from northern Minnesota. As the U.S. steel industry has been increasingly relying on foreign sources of iron ore, the transportation cost advantage of a southern Great Lakes location has been reduced (Hogan, 1987; and Rubenstein, 2014).
Motor vehicle assembly is an example of a bulk-gaining industry. An assembled motor vehicle occupies a much greater volume and is more expensive to ship than the sum of its individual parts. Consequently, carmakers have selected assembly plant sites that minimize their costs of shipping finished vehicles to dealerships. It should be noted, however, that other factors, such as economies of scale in production and the ability to acquire and assemble pieces for a large enough tract of land at an affordable price, also affect carmakers’ plant location decisions (Rubenstein, 1992).

The literature on production agglomeration and industrial location includes numerous papers that estimate the role of specific factors in explaining the distribution of manufacturing in general, as well as that of individual industries (for a detailed tabulation, see Arauzo-Carod, Liviano-Solis, and Manjón-Antolín, 2010). As one of the largest manufacturing industries, the auto industry has been the focus of much attention in this literature. Nearly all of the empirical analysis relies on U.S. and North American data. For instance, Woodward (1992) examines foreign direct investment in the U.S. manufacturing sector, and finds evidence that access to interstate highways plays a crucial role in where foreign carmakers decide to establish plants outside metropolitan areas. Smith and Florida (1994) estimate a model for Japanese-affiliated automotive-related plants within the United States, and find proximity to assembly plants to be an important factor in the location decisions of components supplier plants. Klier and McMillen (2008) estimate the location pattern of motor vehicle parts plants by way of a conditional logit model, and find the observed location choices are well explained by factors linked to agglomeration. Two of these factors—namely, good highway access and shorter distance to assembly plants—suggest that cost considerations of transporting goods play a significant role in supplier plant siting decisions. A third factor—shorter distance to Detroit—suggests such decisions may be partly based on wanting to have ready access to the news and innovations (knowledge spillovers) emanating from the center of the auto industry in the United States. Incidentally, that article illustrates the reorientation of supplier plant locations along the north–south auto alley (from a concentrated area surrounding Detroit, with southern Michigan, Indiana, and Ohio as its hub). It establishes that the degree of agglomeration for old and new plants is quite similar, despite this reorientation. More generally, Klier and Rubenstein (2008, 2013a) provide an extensive account of how the footprint of the auto industry in North America has evolved over the past several decades.

The evolution of North America’s auto production footprint

While motor vehicle production in North America tends to be concentrated today, as it has in the past, the industry’s footprint has changed over time (this section draws heavily on Rubenstein, 1992). In North America, motor vehicle production experienced two periods of intense agglomeration. The first period began during the first decade of the twentieth century, and the second started during the 1980s. In the intervening years, the industry’s assembly footprint broadened with the establishment of the system of branch assembly plants.

When commercial production of motor vehicles began in the United States during the 1890s, more than half of the firms were located in the Northeast. The distribution of producers changed twice in the early twentieth century. First, most production shifted from the Northeast to southeastern Michigan. Motor vehicle production clustered in southeastern Michigan in the first years of the twentieth century primarily because of existing agglomerations of southeastern Michigan with southern Michigan, Indiana, and Ohio as its hub). It establishes that the degree of agglomeration for old and new plants is quite similar, despite this reorientation. More generally, Klier and Rubenstein (2008, 2013a) provide an extensive account of how the footprint of the auto industry in North America has evolved over the past several decades.

The evolution of North America’s auto production footprint

While motor vehicle production in North America tends to be concentrated today, as it has in the past, the industry’s footprint has changed over time (this section draws heavily on Rubenstein, 1992). In North America, motor vehicle production experienced two periods of intense agglomeration. The first period began during the first decade of the twentieth century, and the second started during the 1980s. In the intervening years, the industry’s assembly footprint broadened with the establishment of the system of branch assembly plants.

When commercial production of motor vehicles began in the United States during the 1890s, more than half of the firms were located in the Northeast. The distribution of producers changed twice in the early twentieth century. First, most production shifted from the Northeast to southeastern Michigan. Motor vehicle production clustered in southeastern Michigan in the first years of the twentieth century primarily because of existing agglomerations of southeastern Michigan with southern Michigan, Indiana, and Ohio as its hub). It establishes that the degree of agglomeration for old and new plants is quite similar, despite this reorientation. More generally, Klier and Rubenstein (2008, 2013a) provide an extensive account of how the footprint of the auto industry in North America has evolved over the past several decades.
Carmakers began to abandon the branch assembly plant system during the 1960s. As a result, the assembly footprint began to tighten again. The impetus for this change was the introduction of a variety of models of different sizes that could not be built on the same assembly line. Instead, each assembly plant was devoted to producing models of a single size. Consistent with the principles of industrial location theory based on Weber (1929), carmakers calculated that if a single-sized model was to be assembled at a single plant, its location needed to minimize the aggregate cost of shipping to consumers throughout North America. That region became known as auto alley (see figure 1 on p. 102). In calculating the optimal location for their assembly plants, carmakers determined the key factor was the percentage of the nation’s dealerships that could be reached within a one-day drive by truckers. (From the 1960s onward, trucks were the primary mode for delivering assembled vehicles.) A location in auto alley makes possible one-day delivery to population centers between New York and Texas. Note that auto alley is several hundred miles east of the U.S. center of population, located in Missouri—which is too far west to permit one-day delivery to the East Coast markets and not far enough west to reach California markets in one day.

Later on, in the 1980s, Japanese-owned carmakers began to assemble vehicles in the United States. They too determined that in order to minimize the costs of shipping their final products throughout North America, the optimal locations for their assembly plants were in auto alley (Woodward, 1992).

The history of Europe’s fragmented auto production footprint

In this section, we discuss the historical distribution of motor vehicle production in Europe. In the following section, we highlight factors that precipitated changes in this distribution late in the twentieth century.

For most of the twentieth century, motor vehicle production in Europe was fragmented among the region’s individual countries. Europe’s automobile industry represented a collection of national industries, with each of the major vehicle-producing countries dominated by one or a small number of vehicle producers headquartered in that country—the so-called national champions (Lagendijk, 1997). Fiat of Italy, Volkswagen (VW) of Germany, Renault and Peugeot of France, and Leyland and Rover of the United Kingdom were examples of national champions during the twentieth century. While consumers were able to choose among products from virtually every auto producer in Europe, each vehicle assembler’s production facilities were clustered in the country where it was headquartered (see figure 5, which illustrates this pattern for 1990). Accordingly, most motor vehicles sold in France were produced in France by French companies, and most motor vehicles sold in Italy were produced in Italy by Italian companies—and this pattern also held in the United Kingdom, Spain, and less populous countries, such as Sweden and Czechoslovakia. While the U.S. auto industry clustered in the early twentieth century around the single geographical node of southeastern Michigan, Europe featured multiple nodes, including Paris, France; Wolfsburg, Germany; Turin, Italy; and Coventry, UK (Bentley, Bailey, and MacNeill, 2013; and Lung, 2004).

Despite the polycentric nature of Europe’s motor vehicle industry, which persisted for most of the twentieth century, the region’s pioneering carmakers were collaborating across national boundaries as early as the late nineteenth century. For example, engines from Germany’s Benz (a predecessor of today’s Daimler) were used in vehicles made throughout Europe, and France’s Panhard et Levassor (a predecessor of today’s PSA Peugeot Citroën) created the assembly platform structure with the engine in front that became the auto industry’s standard (Flink, 1988, pp. 15–22). But during World Wars I and II and the interwar period, protectionist policies effectively precluded region-wide agglomeration of Europe’s motor vehicle industry (Flink, 1988, p. 253). That is, each of the more populous countries in Europe protected its own motor vehicle industry from producers of other countries through high tariffs—which deterred region-wide auto production agglomeration. In 1915, for example, the United Kingdom imposed a 25 percent tariff on the import of assembled vehicles, as well as a 10 percent tariff on imported parts (Womack, Jones, and Roos, 1990, p. 228). In 1931, the average tariff on the import of manufactured goods was 30 percent in France, 21 percent in Germany, 46 percent in Italy, and 63 percent in Spain (Bairoch, 1993, p. 40). As a result of tariffs, as well as patriotic loyalty to domestic brands, the distribution of motor vehicle production across Europe remained polycentric. The British auto industry agglomerated around Coventry in the United Kingdom, the Italian auto industry around Turin in Italy, and so on across Europe’s individual countries.

Spatial fragmentation of Europe’s auto production persisted after World War II. The Iron Curtain divided Germany in two and isolated Central European countries from most consumer goods produced in Western Europe. The agreements and organizations promoting economic unity in Western Europe, beginning with the
1951 Treaty of Paris and 1957 Treaty of Rome, removed some of the barriers to trade within the six original members of the European Economic Community, or EEC (Belgium, France, Italy, Luxembourg, the Netherlands, and West Germany). However, some restrictions on the trade of motor vehicles across borders remained in place (Flink, 1988, pp. 298–299).

Incidentally, agglomeration of motor vehicle production was also discouraged by EEC policies promoting investment in areas within the six member states—which were suffering from high levels of poverty and unemployment, as well as significant losses in manufacturing. As part of these policies, the manufacturing sector received incentives to disperse new production facilities to areas not traditionally associated with industrial production, such as Belgium, southern Italy, and western France (Lung, 2004).

The economic slowdown in the 1970s, triggered by the two oil shocks of the decade, brought some restructuring to motor vehicle production in Europe. In response to the slowdown, some countries protected their national champions of automotive production and some did not. France and Italy were examples of the former: They maintained policies protecting their domestic carmakers, thereby limiting the entry of foreign competitors, especially those based in Asia. Additionally, domestic carmakers based in France and Italy consolidated around this time: In Italy, Fiat took over Lancia (in 1969) and Alfa Romeo (in 1986); and in France, Peugeot merged with Citroën (in 1976) (Flink, 1988).

Meanwhile, major changes in motor vehicle production were occurring in the United Kingdom and Spain. In the UK, after the consolidation of the domestic
auto industry in the 1960s, the last British-owned carmaker, British Leyland (later renamed Rover), still struggled in the 1970s and 1980s (despite receiving government aid). Eventually, its assets, including its UK plants, were sold to foreign producers (Gibbs, 2013; and Northedge, 2009). In Spain, foreign carmakers built a number of assembly plants, in particular during the 1970s and 1980s. At the time, Spain was considered an attractive location for the combination of its proximity to European consumers and its relatively low labor costs within Western Europe (Pallares-Barbera, 1998).

Figure 5 (on p. 107) shows the distribution of auto assembly plants in Europe in 1990. The proximity of assembly plants to each other should not be mistaken for agglomeration influenced by economic factors. The locations of the three largest auto assembly plants in Europe circa 1990—Wolfsburg, Germany; Sochaux, France; and Mirafiori, Italy—illustrate the lack of importance of economic geography factors in site selection (as discussed earlier in the section on agglomeration and industrial location theories). Wolfsburg, a city constructed from scratch by the Nazis during the 1930s in order to build “the people’s car,” is located in Lower Saxony, a state of Germany not traditionally associated with manufacturing, some distance away from the Ruhr Valley, historically the center of production (and population) in Germany (Flink, 1988). Sochaux—in far eastern France, fairly remote from Paris, the country’s principal market—was the home of the Peugeot family (Flink, 1988). In Italy, the Mirafiori factory was built in a suburb of Turin—hometown of Fiat’s owners, the Agnelli family, but quite far north of Rome, the country’s largest market (Clark, 2012).

Toward auto production agglomeration across Europe

A sizable literature describes and analyzes the evolution of the auto industry in Europe (see, for example, Bentley, Bailey, and MacNeill, 2013, for an extensive discussion, as well as Domański and Lung, 2009; Domański and Gwosdz, 2009; Jürgens and Krzywdzinski, 2009; and Lung, 2004). During the early 1990s, two key economic and political events had a profound impact on Europe’s motor vehicle industry: the ratification of the Treaty on European Union (the Maastricht Treaty) and the dismantling of the Iron Curtain. These two events enlarged the market of Europe and ultimately enabled a more agglomerated distribution of motor vehicle production across the entire continent.

The Treaty on European Union—the single most important step in European unification—was drafted in 1991, signed by the 12 member countries in 1992, and implemented beginning in 1993. It established common foreign and security policies; assigned stronger decision-making authority to the European Parliament and other supranational institutions; and set the groundwork for the economic and monetary union, including the introduction of the euro as a common currency across the region. For the auto industry, European unification ushered in a uniform region-wide regulatory framework for energy efficiency and pollution-control technologies, as well as other vehicular technologies. Tariffs were removed, and national differences in the pricing of motor vehicles were reduced (Lung, 2004).

The dismantling of the Iron Curtain unified the western and eastern parts of Europe, and rather quickly led to a considerable enlargement of the market for motor vehicles and other goods. The fall of communism in Europe that occurred between 1989 and 1991 brought about democratically elected governments in several countries that are now referred to as Central Europe (see p. 102 for the countries we consider to be part of this region for the purposes of this article). Negotiations between certain Central European countries and the European Union removed trade barriers during the 1990s, and culminated in the accession into the European Union of Czechia, Hungary, Poland, Slovakia, and Slovenia in 2004 and of Bulgaria and Romania in 2007. EU membership required the adoption of EU regulations—a source of stability for the conduct of business in Central Europe. As a result, making a substantial investment in Central Europe became not only a feasible option but an attractive one for carmakers to consider (Domański, Klier, and Rubenstein, 2014; and Domański and Lung, 2009).

More than any other factor, the accession of Central European countries into the European Union promoted region-wide integration of the region’s motor vehicle industry. Central Europe was home to an auto industry under communism. Yet under communist rule, factories there produced small cars and trucks that were quite different from those in Western Europe and that had little appeal to consumers in other countries. Trade barriers to the West had resulted in the import (export) of very few new vehicles into (out of) Central Europe. By contrast, in a unified Europe, motor vehicles no longer varied widely among the region’s individual countries. Automobile production subsequently expanded eastward in a major way.

This expansion of the auto industry’s footprint led to a significant increase in its production capacity. As Central Europe increased its share of Europe’s light vehicle production from just over 5 percent in 1990 to 20 percent within two decades (Domański, Klier, and Rubenstein, 2014), more than ten new assembly
plants were opened in that region (see figure 6). The auto assemblers’ move into Central European countries was motivated by a desire to gain access to the newly opened local markets, as well as by the cost advantages (especially for labor) these nations offered (Domański and Lung, 2009). Between 1990 and 2013 the number of large assembly plants (capable of producing at least 100,000 vehicles per year) in Central Europe rose from eight to 18. The majority of these additional plants were located in just three countries: Czechia, Poland, and Slovakia. Prior to that time, Spain and Portugal had been the main beneficiaries of new automobile investment in Europe.

Around 1990, European carmakers also faced two key challenges that threatened their long-term competitive positions in their respective home markets. One principal challenge concerned quality and productivity issues. The 1990 book *The Machine That Changed the World*—produced by researchers with the International Motor Vehicle Program (IMVP) at the Massachusetts Institute of Technology (MIT)—was highly critical of European motor vehicle industry practices, suggesting that European producers lagged their international competitors in terms of productivity and quality. The other principal challenge to European carmakers at the time was increased competition from Japanese carmakers. In 1990 Japanese carmakers had a more modest presence in Western Europe as a whole than in North America (where they had opened seven assembly plants during the 1980s). Japanese auto companies held 24 percent of light vehicle sales in the United States in 1990; in contrast, Japanese auto assembly firms had only 11.6 percent of comparable sales in Western Europe in 1990.
plants in Europe in 1990 were operated by Nissan in Sunderland, UK, and in Barcelona, Spain; and neither had yet reached annual production levels of at least 100,000 units. Conventional wisdom in 1990 was that Japanese carmakers would soon attain in Europe as a whole the levels of sales and production found in North America and in smaller European countries (Dicken, 1992; and Lagendijk, 1997). European carmakers responded to both challenges by adopting Japanese-inspired production methods that closed the gaps in quality and productivity with their foreign competitors. More to the point of this article, European carmakers also altered the spatial distribution of auto production within Europe, favoring locations consistent with Weber’s (1929) principles of factory site selection for an industry producing bulk-gaining goods. This was yet another way for European auto assemblers to improve their productivity in response to competition from North America and Asia.

**Features of agglomerated auto production in Europe**

As we explained in detail in the previous section, the changes in the underlying economic geography of Europe encouraged agglomeration of the region’s motor vehicle production. Following the collapse of communism in Central Europe, the footprint of Europe’s motor vehicle production changed from a multinational polycentric distribution (with production generally self-contained within individual countries) to a region-wide agglomeration in a corridor with a northwest–southeast axis (figure 6 on p. 109). The new distribution of auto production across Europe more closely resembles the clustering of auto production observed in North America. Auto producers in Europe today are optimizing their operations over a much larger area than before 1990, choosing plant locations that minimize the costs of reaching a large market. In this section, we elaborate on two distinctive trends of Europe’s increasingly agglomerated distribution of motor vehicle production: namely, 1) that most new auto production facilities have been situated inside an agglomerated corridor and 2) that new automotive investment within the corridor has been occurring primarily in the eastern portion.

**Changing distribution of auto production facilities across the whole of Europe**

In 1990, carmakers operated 74 large assembly plants in Europe (with each plant capable of producing at least 100,000 vehicles per year). Between 1990 and 2013, 20 large auto assembly plants were opened in Europe and 14 were closed. As a result of these changes, in 2013 Europe had a total of 80 large auto assembly plants (see table 1).

Europe’s large auto assembly plant count had risen since 1990 mostly because of the clustering of investment in the region’s auto corridor. The number of large plants inside the corridor increased from 52 in 1990 to 58 in 2013, while the number outside the corridor remained at 22. Forty-one of the 52 large plants inside the corridor in 1990 were still open in 2013, while 11 were closed and 17 new ones were opened (see figure 7 and table 1). Meanwhile, 19 of the 22 large plants outside the corridor in 1990 were still open in 2013, while three were closed and three new ones replaced them (see figure 7 and table 1).

As a result of these plant openings and closures, the percentage of Europe’s large auto assembly plant production located in the corridor increased noticeably between 1990 and 2013: In 1990, 68 percent of the region’s 14.4 million vehicles were assembled in the corridor; however, in 2013, 78 percent of the region’s 15.4 million vehicles were assembled in the corridor (see table 1). By comparison, approximately 73 percent of North American production took place in auto alley in 2013.

As a bulk-gaining industry, the motor vehicle industry tends to have its assembly plants agglomerate in order to minimize the costs of shipping the final products to the consumers. Consistent with this principle of industrial location theory based on Weber (1929), Europe’s auto production corridor is situated within the continent’s area of highest population concentration. Figure 8 shows each NUTS-3 region’s level of population within a 450-kilometer radius from its centroid.

![Figure 8](image-url)
as of 2013. The highest values appear for NUTS-3 regions within Germany, northeastern France, and western Czechia. This area represents the heart of Europe’s auto production corridor.

Within a one-day drive of Europe’s auto production corridor (roughly 600 kilometers) are clustered approximately 80 percent of the region’s population, about 85 percent of the region’s total gross national income (and therefore buying power for new vehicles), and around 70 percent of the region’s new vehicle sales in 2014.

As the region’s largest auto producer, Volkswagen has been the carmaker with the most location decisions affecting (and affected by) the agglomeration of assembly plant production in Europe. VW built six of the 20 new plants in Europe between 1990 and 2013 and owned only one of the 14 that were closed (a plant in Barcelona, Spain, that was replaced with a new facility nearby). In addition, VW took over five plants in Central Europe—specifically, three in Czechia and one each in Slovakia and Poland—along with one plant each in the former East Germany, Portugal, and Spain. Asian carmakers opened seven new plants across Europe—specifically, two each in Czechia and the UK and one each in France, Hungary, and Slovakia (two of the plants were joint ventures between Toyota and PSA).

Most auto parts suppliers’ plants are also located in the auto production corridor, although the percentage is lower than that for large auto assembly plants. The 100 largest parts suppliers (by revenues) together had 1,825 plants in Europe in 2011, and 74 percent of these were located in the auto production corridor.

Moreover, of the nearly 1 million employees working in the auto production corridor...
for parts suppliers in 2010, about 688,000, or around 70 percent, of them had their auto parts jobs in the corridor (see table 2).

**Assembly plant distribution within the auto production corridor**

Europe’s auto production corridor is subdivided into a western portion and an eastern portion. The dividing line follows the old Iron Curtain. The eastern portion encompasses the auto assembly plants in the Central European countries of Czechia, Hungary, Poland, and Slovakia, as well as the former East Germany. The western portion encompasses the auto assembly plants in Austria, Belgium, France, the Netherlands, the United Kingdom, and the former West Germany. The western portion accounted for 56 percent of all large auto assembly plant production in Europe in 2013 (down somewhat from 64 percent in 1990), while the eastern portion accounted for 22 percent of it (up substantially from 4 percent in 1990) (see table 3).

Most of the new assembly plants were sited in the eastern portion of the auto production corridor, while most of the closed plants were situated in the western portion (see figure 6 on p. 109 and figure 7 on p. 111). In 1990, the western portion had 44 of the corridor’s

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Location</th>
<th>Number of jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western portion of corridor</td>
<td>390,299</td>
</tr>
<tr>
<td>Germany</td>
<td>244,382</td>
</tr>
<tr>
<td>France</td>
<td>61,759</td>
</tr>
<tr>
<td>United Kingdom (estimated)</td>
<td>56,600</td>
</tr>
<tr>
<td>Belgium</td>
<td>10,947</td>
</tr>
<tr>
<td>Austria</td>
<td>11,956</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4,655</td>
</tr>
<tr>
<td><strong>Eastern portion of corridor</strong></td>
<td><strong>297,054</strong></td>
</tr>
<tr>
<td>Poland</td>
<td>105,762</td>
</tr>
<tr>
<td>Czechia</td>
<td>102,425</td>
</tr>
<tr>
<td>Hungary</td>
<td>51,617</td>
</tr>
<tr>
<td>Slovakia</td>
<td>37,250</td>
</tr>
<tr>
<td><strong>Outside corridor</strong></td>
<td><strong>295,911</strong></td>
</tr>
<tr>
<td>Romania</td>
<td>97,072</td>
</tr>
<tr>
<td>Italy</td>
<td>86,022</td>
</tr>
<tr>
<td>Spain</td>
<td>66,421</td>
</tr>
<tr>
<td>Portugal</td>
<td>21,433</td>
</tr>
<tr>
<td>Sweden</td>
<td>16,454</td>
</tr>
<tr>
<td>Slovenia</td>
<td>8,509</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>983,264</strong></td>
</tr>
</tbody>
</table>

Source: Authors’ calculations based on data from Frigant and Miollan (2014).
52 large auto assembly plants (capable of producing 100,000 vehicles per year) and the eastern portion had eight. Twelve of the 17 large assembly plants that opened in the corridor between 1990 and 2013 were in the eastern portion. In contrast, nine of the 11 large auto assembly plant closures in the corridor over this period were in the western portion. As a result, the number of large assembly plants in the eastern portion increased from eight in 1990 to 18 in 2013 (specifically, from three to five in Czechia, from zero to three in Hungary, from one to three in Slovakia, and from one to four in the former East Germany). Meanwhile, the number of large assembly plants in the western portion of the corridor declined from 44 to 40; three of the four net closures occurred in Belgium, while very few changes in the number of large plants occurred in the three primary car-producing countries in the western portion (that is, the number of assembly plants remained at 15 in the former West Germany and at 11 in France, but decreased from 11 to ten in the UK) (see table 3). The corresponding increase of the auto corridor’s production share (up from 68 percent in 1990 to 78 percent in 2013) was even more tilted in favor its eastern portion: The light vehicle production share of the western part of the corridor dropped by 8 percentage points—from 64 percent to 56 percent—whereas the eastern part’s share more than quintupled—from 4 percent to 22 percent (see table 3).

As indicated earlier, the carmakers’ push into Central Europe after the dismantling of communism and removal of trade barriers was initially motivated by the lower costs of labor in the countries there. The hourly compensation cost for auto workers in 2007, for example, was (at a purchasing power parity, or PPP, exchange rate$^{33}$) $11.30 in Poland, $12.30 in Hungary, and $16.64 in Czechia, compared with $44.47 in Germany, $26.34 in France, $28.02 in Spain, $27.66 in the UK, and $24.64 in Italy (Stanford, 2010, p. 392).

Several indicators suggest that auto production has become integrated across a much wider area of Europe than it was in 1990. First, the tight linkage between motor vehicle production by the national champions (VW, Fiat, PSA, and Renault–Nissan$^{34}$) and their respective home countries has been noticeably eroded (see figure 9 and table 4). Today all four of these major European auto producers have a substantial production presence in Central Europe. Their new investments in production facilities sited in Central European countries lowered the share of auto production in their respective home countries (to just around 50 percent of total European auto production in 2013 from well above the 50 percent mark in 1990; see table 4).$^{35}$ For further evidence of motor vehicle production having become integrated across a wider expanse of Europe than in 1990, we point to two low-cost automotive brands that are produced in Central Europe for sale in all of Europe. Renault–Nissan produces Dacia vehicles primarily at its Pitești plant in Romania. VW produces Škoda vehicles primarily at two plants in Czechia (Mladá Boleslav and Kvasiny). For both brands, most of their European sales now take place in Western Europe. For Dacia, the Western European sales share grew from 0 percent to 83 percent between 2000 and 2013; for Škoda, it increased from 55 percent to 74 percent over the same period.$^{37}$

**TABLE 3**

<table>
<thead>
<tr>
<th></th>
<th>1990</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Western of corridor</td>
<td>Eastern of corridor</td>
</tr>
<tr>
<td>Number of large assembly plants</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>Percentage of large assembly plants</td>
<td>59</td>
<td>11</td>
</tr>
<tr>
<td>Percentage of production</td>
<td>64</td>
<td>4</td>
</tr>
</tbody>
</table>

Notes: The corridor refers to the auto production corridor in Europe. See table 2 for a list of the countries in each of the three geographical categories. In this table, Germany is accounted for somewhat differently than it is in tables 1 and 2: Plants located in the former West Germany are included in the western portion of the auto corridor, while plants located in the former East Germany are included in the eastern portion of the auto corridor, as noted in the text. This is done to allow for a more detailed discussion of the geography within the auto corridor. The sample for this table is limited to large auto assembly plants, which are defined as plants capable of producing at least 100,000 units per year. Certain percentage rows may not total because of rounding.

Sources: Authors’ calculations based on data from ACEA, IHS Global Insight, and auto company websites.
Outliers of auto production agglomeration in Europe and North America

From our previous discussion and map figures, it should be clear that motor vehicle production in Europe and North America does not take place exclusively inside their respective auto industry corridors today. Some investment continues to flow toward production facilities located outside these corridors. In this section, we briefly touch on the principal exceptions to the patterns of agglomeration of motor vehicle production in Europe and North America.

In Europe, 19 of the 22 large plants in operation outside the corridor in 1990 remained open in 2013, while three were closed and three new ones were opened (see table 1 on p. 110). In both 1990 and 2013, Spain had ten of the 22 assembly plants and Italy had seven. The three closed plants were in Italy, Spain, and Sweden, and the three new ones were in Italy, Portugal, and Spain. In Spain, the new plant in the Barcelona area was built to replace the one that had been shuttered there. Yet, simply reporting the counts of auto assembly plants in 1990 versus 2013 can be a bit misleading because production volumes within assembly plants can be adjusted noticeably. The percentage breakdown of European auto output by location shows a larger change in the production footprint in Europe: For plants outside the auto corridor, the share of production declined noticeably from 32 percent in 1990 to 22 percent in 2013 (see table 1 on p. 110).

In Spain, large-scale investment in motor vehicle production began in the early 1950s. A Spanish-owned carmaker SEAT (Sociedad Española de Automóviles...
de Turismo) was established in 1950 by a state-owned industrial holding company, and it became a subsidiary of VW in 1990. In addition to SEAT’s Barcelona plant, four other assembly plants were opened in Spain in the 1950s, and these four are still in use; they were built by Daimler, Renault, Chrysler, and Citroën (the plants built by the last two are now owned by PSA). A second wave of investment in the late 1970s and early 1980s resulted in the openings of five more of Spain’s current inventory of ten assembly plants. Carmakers are expected to maintain the current roster of assembly plants in Spain, but new assembly plants have been situated by Renault–Nissan and PSA on the other side of the Strait of Gibraltar, in Morocco (to take advantage of relatively lower labor costs there). New vehicle production in Spain stood at around 2 million units in both 1990 and 2013.

Since 1990, auto production in Europe has been increasing in an extension of the corridor toward the east. Romania, which we did not include in the definition of the corridor, is a major production center for Renault–Nissan. Several carmakers have important production facilities in Turkey that export vehicles to Europe. After the breakup of the Soviet Union, production in Russia also appeared likely to become integrated into the European market, but events in recent years have pushed that likelihood into a future beyond the planning horizon for carmakers’ investment decisions (see, for example, Adamanis, 2015).

In North America, the principal center of motor vehicle production outside auto alley is in central Mexico. Sixteen assembly plants were in operation in 2013, with three more under construction at that time. Production in Mexico increased from 820,558 vehicles in 1990 to 3,038,196 in 2013, so not surprisingly, its share of North American auto production rose from 6 percent in 1990 to 19 percent in 2013. The primary factors driving increased auto production in Mexico have been relatively low production costs and strong export opportunities (Klier and Rubenstein, 2013b). Approximately 60 percent of Mexico’s auto production is for export to the United States and Canada, about 20 percent is for export to other regions, and around 20 percent is for domestic consumption.19

### Conclusion

The footprints of motor vehicle assembly (and parts) production in North America and Europe are concentrated today. In this article, we laid out in broad strokes the process by which the geography of Europe’s light vehicle industry has changed since 1990. It turns out that the auto industry in Europe restructured its footprint according to the principles of agglomeration and industrial location theories (including the key principle of locating assembly plants closer to customers for bulk-gaining industries). Those principles also help explain the current footprint of motor vehicle assembly observed in North America. Today both North America and Europe display agglomerated distributions of motor vehicle production that are strikingly similar. While the two regions have reached their respective patterns of auto production agglomeration based on different histories, similar forces have shaped the spatial convergence of auto production now observed in both regions.

North America’s spatial pattern of auto production has been formed through decisions made by international carmakers to site their plants in auto alley and decisions by Ford, General Motors, and Chrysler to replace coastal branch assembly plants with new plants in auto alley. Europe’s auto production corridor has developed primarily through decisions by the region’s long-standing national champions—VW, Fiat, PSA, and Renault–Nissan—to replace country-specific production strategies designed to serve country-specific markets with region-wide production strategies designed to serve the economic space of an enlarged Europe following the fall of the Iron Curtain.
In this section, unless indicated otherwise, references to assembly and parts plant counts and light vehicle production (and their percentages by various geographical divisions) in Europe are from authors’ calculations based on data obtained from the ACEA (Association des Constructeurs Européens d’Automobiles), or, in English, the European Automobile Manufacturers’ Association; see http://www.acea.be, IHS Global Insight, and auto (assembly and parts) company websites. Also, unless indicated otherwise, references to assembly and parts plant counts and light vehicle production (and their percentages by various geographical divisions) in North America are from authors’ calculations based on data from Ward’s AutoInfoBank (database by subscription), Elm Analytics, and auto company websites.

1For a description of the EU and further details on its history and policies, see http://europa.eu/index_en.htm.

4The Czech Republic is the English translation of the country’s constitutional name, but Czechia is the official English short form of the name according to the United Nations. The use of Czechia instead of the Czech Republic is equivalent to the use of the United Kingdom instead of the United Kingdom of Great Britain and Northern Ireland. See http://www.czechia-initiative.com.

23As shown in figure 1 (p. 102), auto alley is approximately 1,300 kilometers long and 400 kilometers wide.

In figures 2 and 5–9, we refrain from labeling individual countries (or states) because the plant data would be obscured. For a detailed map of Europe with all the countries identified, use the map function at http://europa.eu/about-eu/countries/index_en.htm. Moreover, the set of auto assembly plants shown in these map figures is slightly less restrictive than the one used in tables 1 and 3, where we limit ourselves to just large plants. Large plants are defined as plants capable of producing at least 100,000 units per year.

3According to Google Maps, from Stuttgart, Germany, it is 11 hours by truck to Coventry, United Kingdom; nine hours to Budapest, Hungary; 12 hours to Barcelona, Spain; and 13 hours to Naples, Italy. We include the United Kingdom in the European auto production corridor based on road travel times.

4Analysis of European auto parts suppliers is based on data obtained by authors from the websites of the 100 largest parts suppliers (by revenues) in Europe. The data referenced here apply to 2011. (For an earlier version of these data [for 2010], see Klier and McMillen, 2013.)

5In theory, industry clustering leads to labor market pooling, facilitating economies of scale. As firms fail, workers can quickly move to other employers, thereby maximizing their productivity and lowering the variance in worker wages. Moreover, clustering facilitates better worker–firm matches.

6In the past, each assembly plant produced a much smaller number of vehicles than the 200,000 per year that are typically made today. For example, in 1925 Ford assembled 1.5 million vehicles at 32 assembly plants, for an annual average of fewer than 50,000 vehicles per plant (Rubenstein, 1992).

7The location of parts-producing plants has shifted southward (from the area centered on southeastern Michigan) as well. However, as noted before, Klier and McMillen (2008) find that the degree of agglomeration among parts supplier plants has remained essentially unchanged.

11In this section, unless indicated otherwise, references to assembly and parts plant counts and light vehicle production (and their percentages by various geographical divisions) in Europe are from authors’ calculations based on data from the ACEA (Association des Constructeurs Européens d’Automobiles, or, in English, the European Automobile Manufacturers’ Association; see http://www.acea.be), IHS Global Insight, and auto (assembly and parts) company websites. Also, unless indicated otherwise, references to assembly and parts plant counts and light vehicle production (and their percentages by various geographical divisions) in North America are from authors’ calculations based on data from Ward’s AutoInfoBank (database by subscription), Elm Analytics, and auto company websites.

The official company name is PSA Peugeot Citroën, but it is commonly referred to as PSA.

8Over the period 1946–90, the term Iron Curtain referred to the border between the democratic countries of Western Europe and the communist countries of Central and Eastern Europe (the Eastern Bloc). For more details, see http://www.britannica.com/event/Iron-Curtain.

9The European Economic Community was informally known as the Common Market. The European Union replaced the EEC in 1993. For more details, see http://www.britannica.com/topic/European-Community-European-economic-association.

10Spain and, to some extent, Portugal had managed to attract initial auto assembly plant investment long before then, during the 1950s, because foreign auto producers wanted to gain footholds in those two markets. The presence of auto assembly operations in southwestern Europe grew noticeably during the second half of the 1970s and the early 1980s, before both countries joined the European Union in 1986.

11Note that during the Cold War (1947–91) many Central European countries, such as Poland, Czechoslovakia, Romania, and East Germany, had their own automobile industries and motor-vehicle-producing companies. As Central (and Eastern) Europe opened up to market-based competition and outside competitors, those motor vehicle producers either went under or were absorbed by Western European producers (for example, Volkswagen acquired the Czech producer Škoda, and Renault acquired the Romanian producer Dacia). In addition, many new auto assembly operations (so-called greenfield plants, such as Opel’s plant in Eisenach, which is within the former East Germany) were established (information from IHS Global Insight and auto company websites).

12Domański and Lung (2009, p. 5) state that “on the whole, the relatively high productivity of the CEE [Central and Eastern Europe] labour force based on its skills and motivation together with the new technology and organization of production and combined with lower wages led to the attraction of a great amount of foreign direct investment in the auto industry during the last 15 years.”

13An additional plant was announced in 2014. It is to be built in Wrezinia, Poland. Production at the plant is scheduled to start in 2016. See http://www.volkswagen-poznan.pl/en/new-plant.

14Authors’ calculations based on data from ACEA, IHS Global Insight, and auto company websites.


16Authors’ calculations based on data from Ward’s AutoInfoBank and Crain Communications Inc. (1992).
IHS Global Insight.

For an example of how consumers have responded to improvements in European automotive quality, see J.D. Power (2014).

Note the differences in the footprint of auto production in 2013 (figure 6 on p. 109) versus 1990 (see figure 5 and its description on pp. 106–108).

Authors' calculations based on data from Ward's AutoInfoBank.

NUTS stands for Nomenclature of units for territorial statistics (or, in French, Nomenclature des Unités territoriales statistiques). The NUTS classification is a standard developed and maintained by the European Union for dividing up its members' territories in order to produce regional statistics. The NUTS system favors existing administrative units. For more details, see http://ec.europa.eu/eurostat/web/nuts/overview.

Authors’ calculations based on data from Eurostat.

This information is from IHS Global Insight and auto company websites.

Authors’ calculations based on data obtained from company websites of the 100 largest auto parts suppliers (by revenues).

Within Europe there was a widespread expectation that these new locations of auto production in Central Europe would evolve into sizable centers of new auto demand. See, for example, Jullien and Pardi (2015).

REFERENCES


